DOCUMENT RESUME

ED 104 159 FL 006 816

AUTHOR	Suppes, P.; And Others
TITLE	Developmental Models of a Child's French Syntax. Technical Report No. 243.
INSTITUTION:	Stanford Univ., Calif. Inst. for Mathematical Studies in Social Science.
SPONS AGENCY	National Science Foundation, Washington, D.C.; Office of Naval Research, Arlington, Va.
REPORT NO	TR-243
PUB DATE	Dec 74
NOTE	230p.; Psychology and Education Series; For related document, see ED 076 075
EDRS PRICE	MF-$0.76 HC-$12.05 PLUS POSTAGE
DESCRIPTORS	*Child Language; Developmental Psychology; *French; Generative Grammar; Grammar; *Language Development; *Language Research; Linguistic Theory; Mathematical Linguistics; Models; Psycholinguistics; Research Methodology; Speech; Statistical Analysis; *Syntax; Tables (Data)

ABSTRACT

This is the second report concerned with the analysis of a young child's spoken French. It focuses on the study of the entire corpus of 33 hour sessions occurring approximately once a week and ranging from the time the subject was 25 months old to 38 months old. Chapter 1 is devoted to introductory remarks. Chapter 2 contains a dictionary of the parts of speech used by the subject and summarizes the analysis used on the full corpus. In the third chapter, the grammar developed to account for the subject's speech is presented. In chapter 4 a probabilistic grammar is presented as a more effective way to evaluate the appropriateness of the grammar given in the third chapter. In chapter 5, two developmental models of the child's speech are tested, a stage model and an incremental model. It is concluded that, although neither model fits the data exactly, the evidence in favor of the incremental model is substantial. (Author/PMP)

ED 041159

FL

DEVELOPMENTAL MODELS OF A CHILD'S FRENCH SYNTAX

BY

P. SUPPES, M. LÉVEILLÉ, AND R. L. SMITH

TECHNICAL REPORT NO. 243

DECEMBER 4, 1974

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FL006816

PSYCHOLOGY AND EDUCATION SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

TECHNICAL REPORTS

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

(Place of publication shown in parentheses; if published title is different from title of Technical Report, this is also shown in parentheses.)

125 W. K. Estes. Reinforcement in human learning. December 20, 1967. (In J. Tapp (Ed.), Reinforcement and behavior. New York: Academic Press, 1969. Pp. 63-94.)

126 G. L. Wolford, D. L. Wessel, and W. K. Estes. Further evidence concerning scanning and sampling assumptions of visual detection models. January 31, 1968. (Perception and Psychophysics, 1968, 3, 439-444.)

127 R. C. Atkinson and R. M. Shiffrin. Some speculations on storage and retrieval processes in long-term memory. February 2, 1968. (Psychological Review, 1969, 76, 179-193.)

128 J. Holmgren. Visual detection with imperfect recognition. March 29, 1968. (Perception and Psychophysics, 1968, 4(4), .)

129 L. B. Mlodnosky. The Frostig and the Bender Gestalt as predictors of reading achievement. April 12, 1968.

130 P. Suppes. Some theoretical models for mathematics learning. April 15, 1968. (Journal of Research and Development in Education, 1967, 1, 5-22.)

131 G. M. Olson. Learning and retention in a continuous recognition task. May 15, 1968. (Journal of Experimental Psychology, 1969, 81, 381-384.)

132 R. N. Hartley. An investigation of list types and cues to facilitate initial reading vocabulary acquisition. May 29, 1968. (Psychonomic Science, 1968, 12(b), 251-252; Effects of list types and cues on the learning of word lists. Reading Research Quarterly, 1970, 6(1), 97-121.)

133 P. Suppes. Stimulus-response theory of finite automata. June 19, 1968. (Journal of Mathematical Psychology, 1969, 6, 327-355.)

134 N. Moler and P. Suppes. Quantifier-free axioms for constructive plane geometry. June 20, 1968. (Compositio Mathematica, 1968, 20, 143-152.)

135 W. K. Estes and D. P. Horst. Latency as a function of number of response alternatives in paired-associate learning. July 1, 1968.

136 M. Schlag-Rey and P. Suppes. High-order dimensions in concept identification. July 2, 1968. (Psychometric Science, 1968, 11, 141-142.)

137 R. M. Shiffrin. Search and retrieval processes in long-term memory. August 15, 1968.

138 R. D. Freund, G. R. Loftus, and R. C. Atkinson. Applications of multiprocess models for memory to continuous recognition tasks. December 18, 1968. (Journal of Mathematical Psychology, 1969, 6, 576-594.)

139 R. C. Atkinson. Information delay in human learning. December 18, 1968. (Journal of Verbal Learning and Verbal Behavior, 1969, 8, 507-511.)

140 R. C. Atkinson, J. E. Holmgren, and J. F. Juola. Processing time as influenced by the number of elements in the visual display. March 14, 1969. (Perception and Psychophysics, 1969, 6, 321-326.)

141 P. Suppes, E. F. Loftus, and M. Jerman. Problem-solving on a computer-based teletype. March 25, 1969. (Educational Studies in Mathematics, 1969, 2, 1-15.)

142 P. Suppes and M. Morningstar. Evaluation of three computer-assisted instruction programs. May 2, 1969. (Computer-assisted instruction. Science, 1969, 166, 343-350.)

143 P. Suppes. On the problems of using mathematics in the development of the social sciences. May 12, 1969. (In Mathematics in the social sciences in Australia. Canberra: Australian Government Publishing Service, 1972. Pp. 3-15.)

144 Z. Domotor. Probabilistic relational structures and their applications. May 14, 1969.

145 R. C. Atkinson and T. D. Wickens. Human memory and the concept of reinforcement. May 20, 1969. (In R. Glaser (Ed.), The nature of reinforcement. New York: Academic Press, 1971. Pp. 66-120.)

146 R. J. Titiev. Some model-theoretic results in measurement theory. May 22, 1969. (Measurement structures in classes that are not universally axiomatizable. Journal of Mathematical Psychology, 1972, 9, 200-205.)

147 P. Suppes. Measurement: Problems of theory and application. June 12, 1969. (In Mathematics in the social sciences in Australia. Canberra: Australian Government Publishing Service, 1972. Pp. 613-622.)

148 P. Suppes and C. Ihrke. Accelerated program in elementary-school mathematics--The fourth year. August 7, 1969. (Psychology in the Schools, 1970, 7, 111-126.)

149 D. Rundus and R. C. Atkinson. Rehearsal processes in free recall: A procedure for direct observation. August 12, 1969. (Journal of Verbal Learning and Verbal Behavior, 1970, 9, 99-105.)

150 P. Suppes and S. Feldman. Young children's comprehension of logical connectives. October 15, 1969. (Journal of Experimental Child Psychology, 1971, 12, 304-317.)

151 J. H. Laubsch. An adaptive teaching system for optimal item allocation. November 14, 1969.

152 R. L. Klatzky and R. C. Atkinson. Memory scans based on alternative test stimulus representations. November 25, 1969. (Perception and Psychophysics, 1970, 8, 113-117.)

153 J. E. Holmgren. Response latency as an indicant of information processing in visual search tasks. March 16, 1970.

154 P. Suppes. Probabilistic grammars for natural languages. May 15, 1970. (Synthese, 1970, 11, 111-222.)

155 E. M. Gammon. A syntactical analysis of some first-grade readers. June 22, 1970.

156 K. N. Wexler. An automaton analysis of the learning of a miniature system of Japanese. July 24, 1970.

157 R. C. Atkinson and J. A. Paulson. An approach to the psychology of instruction. August 14, 1970. (Psychological Bulletin, 1972, 78, 49-61.)

158 R. C. Atkinson, J. D. Fletcher, H. C. Chetin, and C. M. Stauffer. Instruction in initial reading under computer control: The Stanford project. August 13, 1970. (In A. Romano and S. Rossi (Eds.), Computers in education. Bari, Italy: Adriatica Editrice, 1971. Pp. 69-99. Republished: Educational Technology Publications, Number 20 in a series, Englewood Cliffs, N. J.)

159 D. J. Rundus. An analysis of rehearsal processes in free recall. August 21, 1970. (Analyses of rehearsal processes in free recall. Journal of Experimental Psychology, 1971, 89, 63-77.)

160 R. L. Klatzky, J. F. Juola, and R. C. Atkinson. Test stimulus representation and experimental context effects in memory scanning. (Journal of Experimental Psychology, 1971, 87, 281-288.)

161 W. A. Rottmayer. A formal theory of perception. November 13, 1970.

162 E. J. F. Loftus. An analysis of the structural variables that determine problem-solving difficulty on a computer-based teletype. December 18, 1970.

163 J. A. Van Campen. Towards the automatic generation of programmed foreign-language instructional materials. January 11, 1971.

164 J. Friend and R. C. Atkinson. Computer-assisted instruction in programming: AID. January 25, 1971.

Table of Contents

Chapter / Section			Page
Acknowledgments			iv
I.	INTRODUCTION		1
II.	DICTIONARY		6
	II.1	Articles	6
	II.2	Nouns	7
	II.3	Pronouns	8
	II.4	Adjectives	10
	II.5	Verbs	11
	II.6	Adverbs	17
	II.7	Prepositions	18
	II.8	Conjunctions of subordination	18
	II.9	Conjunctions of coordination	18
	II.10	Interjections and onomatopeias	18
	II.11	Locutions	19
	II.12	Conclusion	19
III.	GRAMMAR		26
	III.1	Group 1 Highest Level Rules	26
	III.2	Group 2 Incomplete Utterances	28

III.3 Group 3 Noun Phrase and Adjective Phrase Utterances 28

III.4 Group 4 Utterance Combination 32

III.5 Group 5 Utterances with a Verb 36

III.6 Group 6 Nominative Predicate Noun Phrases 39

III.7 Group 7 Subject Noun Phrases 41

III.8 Group 8 Object Noun Phrases 43

III.9 Group 9 Basic Noun Phrases 44

III.10 Group 10 Prepositional Noun Phrases 47

III.11 Group 11 Determiner Introduction 49

III.12 Group 12 Noun-Phrase Utterances 53

III.13 Group 13 Some Pronouns and Common and Proper Nouns 57

III.14 Group 14 Adjective Phrases 58

III.15 Group 15 Post-position Adjectives 59

III.16 Group 16 Determiners 59

III.17 Group 17 Numerical Expressions 60

III.18 Group 18 Verb Phrase Structures 62

III.19 Group 19 Auxiliaries 67

III.20 Group 20 Modals 67

III.21 Group 21 Verbal Groups 68

III.22 Group 22 Verb Phrases with Prepositions or Personal Pronouns 85

III.23 Group 23 Intransitive and Impersonal Verbs . . . 90

III.24 Group 24 Copula 90

III.25 Group 25 Transitive and Pronominal Verbs 90

III.26 Group 26 Predicate Adjectives 91

III.27 Group 27 Adjectives 91

III.28 Group 28 Questions 92

III.29 Group 29 Adverbs of Quantity 94

III.30 Summary results of the grammatical analysis . . . 95

IV. PROBABILISTIC GRAMMAR AND PROBABILISTIC DISAMBIGUATION . . 98

V. DEVELOPMENTAL MODELS 114

V.1 Merging of the data 117

V.2 Reduction of the data 128

V.3 The All-or-None Stage Model 130

V.4 Incremental model 131

V.5 Test of the two-stage model 134

V.6 Fit of the incremental model 137

Appendix I

Incorrect Grammatical Analyses 147

Appendix II

Incorrect Probabilistic Disambiguation 177

References 187

Acknowledgments

Support for this research was provided by the National Science Foundation under grant NSFGJ-443X and the Office of Naval Research under contract N0014-67-0112-0049. This research was performed in collaboration with the Laboratoire de Psychologie Expérimentale et Comparée associé au Centre National de la Recherche Scientifique, Paris, France.

Chapter I

INTRODUCTION

This is the second report concerned with the analysis of a young child's spoken French. The first report [9] was concerned only with the grammar of noun phrases occurring in Philippe's speech. In that first report, in fact, even the noun phrases were restricted to the first three hours, the middle three hours, and the last three hours of the corpus. The details of the collection of the corpus, the recording conditions, and the procedures for transcribing and editing are all described in the first report. We recall here only the fact that the corpus consists of 33 hour sessions occurring approximately once a week and ranging from the time that Philippe was 25 months old to 38 months old.

In the previous report we emphasized constructing and testing a probabilistic grammar. In the present report we move to the entire corpus and are more concerned with the grammar itself and less with its probabilistic aspects, even though a section is devoted to the probabilistic fit of the grammar. We emphasize that we have not abandoned our attitude toward probabilistic grammars outlined in the introduction to the earlier report but simply that, as we moved from the noun phrases to the entire corpus of 33 hours, the detailed development of the grammar itself turned out to be a very complex task, and the presentation of the results of that constructive effort constitutes the largest segment of the present report.

More generally, the present report is organized along the following lines. First, in Chapter 2 we extend the dictionary developed earlier for noun phrases alone to cover the other parts of speech and we present in detail the analysis used on the full corpus. As much as possible we have used classical grammatical categories in constructing the dictionary, and we have followed when possible the kinds of categories used or suggested in Dubois [1].

The long Chapter 3 is concerned with the grammar. Here we are concerned not only to state the generative rules of the grammar but to give illustrative examples of most of the rules in order to provide an intuitive sense of how the rules fit typical instances to be found in the corpus. We emphasize that all of the examples given in the text are actually drawn from Philippe's speech except for one or two cases that are clearly so marked. At the end of the third chapter on the grammar we present summary results of the extent to which the grammar satisfactorily parses the corpus of Philippe's speech. We emphasize the complexity of the grammar we have written for this medium-sized corpus of Philippe's speech. The number of rules in the grammar is 317, and even though this number is large we have not taken into account in the grammar itself, the many inflectional patterns, including plurality of pronouns, verbs etc., even though a full notation for such matters is introduced in the dictionary. In constructing the grammar we have used Dubois, as already mentioned, and also such classical sources as Grévisse [3], but we have found that the

completely systematic approach we have adopted has required us to make emendations and changes on a number of minor matters that only turn up in a full-scale attempt to fit of a grammar to a corpus.

We turn then in Chapter 4 to a brief presentation of the probabilistic grammar, which means that we move from the grammar itself to the introduction of probabilistic parameters for the use of the individual rules. The techniques of analysis used here follow earlier work in the Institute and especially the methodology already applied in the first report on Philippe's speech. In the later part of Chapter 4 we then show how the probabilistic grammar can be used to successfully provide a method of probabilistic disambiguation that we think has interesting applications beyond simply the instances analyzed in detail here. We believe that this method of probabilistic disambiguation represents one of the more interesting applications to the development of probabilistic grammars for a given corpus of speech.

In Chapter 5 we turn to some developmental models of Philippe's speech. Because there is such a natural tendency always to speak in terms of stages of development, we have systematically tried to test a stage model against an alternative incremental model. As might be expected from complex data of the kind collected in a corpus arising from unstructured conversations between Philippe and his parents or other persons, neither model fits the data exactly, but we believe that the methodology is itself of interest and there is sufficient evidence in favor of the incremental model in comparison to the stage model to

challenge the continual use of stage models in conceptual discussions of the linguistic development of children. As far as we know, the present attempt, preliminary though it may be, is among the very first to try a systematic comparison of the two basic kinds of models that can be conceived for the developmental trends in the acquisition of grammatical rules. We would be happier if we were able to present developmental trends for a variety of children, but it will be evident from the details in the present report that the task of testing developmental models in a systematic way for any substantial number of children is an almost overwhelming task at the present level of our technology of data collection and reduction. Our tentative conclusion that the incremental model fits somewhat better than a stage model is of course only applicable to Philippe's speech behavior. We hope that the kind of models we have begun to test in the present report will be of interest to other workers concerned with developmental psycholinguistics and that more systematic comparisons of different types of models of development will be tested on other corpora.

Finally, to avoid any misunderstanding about our use of probabilistic notions in the present report, we want to repeat the essence of the comment we made at the beginning of the earlier report on Philippe's speech. In no sense do we consider the probabilistic account of the use of production rules or the use of probabilistic assumptions for the mean behavior tested by alternative developmental models an ultimate account of Philippe's spoken utterances. It is

apparent enough in a multitude of cases that the probabilities assigned are overruled by the semantics of a particular utterance or by the context in which the utterance occurs. In the case both of the probabilities assigned to the production rules and the averaging assumptions used in the incremental developmental model, it is assumed that the probabilities represent the results of the averaging over a number of utterances, and do not provide a fully detailed model of a particular utterance occurring on a particular occasion. The enormous complexity of the spoken speech of any child, including Philippe, strongly argues for the use of probabilistic methods for an understanding of the central features of the speech, particularly of the central features of development. The present corpus is quite large but the total amount of speech, either spoken or heard, by a young child during the crucial ages from 24 months to 36 months or 42 months is overwhelming in its number and variety. During this critical period we would estimate that Philippe heard on the order of a million words and responded with approximately half as many words. It is our scientific judgment that it will be a very long time before we have developmental models of children's speech sufficiently deep and detailed to avoid the use of probabilistic averaging methods, at least insofar as the objective is to account systematically for a corpus of any size.

Chapter II

DICTIONARY

Philippe I [9] presents in a detailed way the principles which guided the construction of the dictionary. In our present work, which deals with the whole corpus, the dictionary remains the same in outline, but we have added a few subcategories within the 10 main classes of words used in the dictionary (articles, nouns, pronouns, adjectives, verbs, adverbs, prepositions, conjunctions, onomatopeias and interjections, and uncodable expressions). We now present the lexical categories, and discuss the additional subcategories.

II.1 - Articles

This category has not been modified.

Indefinite articles, un, une, are coded IN.
Definite articles, le, la, l', les, are coded DN.

Even though un and une can be considered as cardinal adjectives, it seemed appropriate to classify them unambiguously as articles.

Following Dubois's suggestion [1] (p.152) the words du, de, des, d', au and aux (traditionally thought of as articles) have been coded EP, which stands for prepositions. Such a simplification has the advantage of reducing the lexical ambiguity: had we followed classical grammarians, several of these words should have been multiply

classified (see below II.12). Furthermore, this simplification lets us avoid the delicate question of partitive articles.

II.2 Nouns

Nouns are divided into the two classical categories: common nouns (NC) and proper nouns (NP).

Common nouns are coded as to gender and number.

Common nouns

Gender

A common noun is coded 1 when masculine, 2 when feminine. Homographs (homonyms which have the same pronunciation and spelling) are coded 0. Ex: tour. Words which can be used for both genders are also coded 0. Ex: enfant. Moreover, common nouns which have the same pronunciation for both genders are coded 0. Ex: ami, amie. Eleven words have been coded NC0.0.

Number

In most cases, in spoken French, the plural of common nouns does not differ phonetically from the singular. Consequently, most common nouns have been coded 0. Words for which the plural differs from the singular are coded 1 for singular (ex: cheval, NC1.1) and 2 for plural (ex: animaux, NC1.2). However, Philippe did not always use the correct form of the words; by referring to the context and to the determiner preceding the noun, we were able to transcribe the noun so that its number is clear. For instance, x was omitted when Philippe said chevau when pointing at a horse drawn on a book, and s was added to cheval when Philippe said des chevals. Consequently, chevau is coded NC1.1, while NC1.2 is the code for chevals and oeils.

Proper nouns

Proper nouns are coded to reflect gender only. For example, Papa is coded NP1, Jeanine NP2. Most proper nouns of towns are coded 0, because we found it difficult to decide about their gender, and as Grévisse [3] puts it (p. 207) "in this matter, the usage is very undecided in almost all cases".

II.3 Pronouns

In Philippe I [9], the five main categories of pronouns were personal, demonstrative, possessive, interrogative and relative, and indefinite. The categories have not changed. However, we have entirely modified the coding of personal pronouns.

Personal pronouns are coded by taking into account their position within a subject-verb-object sentence, since it is their position that determines their syntactic form and semantic role.

Consider the following set of sentences extracted from Philippe's speech:

a) Je veux un gâteau.

b) il est pour moi.

c) Je le défais

d) Papa il m'a acheté un bateau avec des roues

e) tu me le défais tout

f) remets-la sur le camion.

g) penses-tu !

h) moi, je vais manger mon croissant

There are three kinds of pronouns in the above sentences.

1) Pronouns that occur before the verb and are the first in the phrase, such as je in a) and c). These pronouns are the subjects of the sentence, and they are coded PF.

2) Pronouns that are in apposition to the subject, such as moi in h), are also coded PF. Pronouns such as moi, nous, vous, lui that occur in two or more positions, as shown in the following examples, are multiply coded.

je vous jure
il lui est arrive des ennuis
vous voulez des legos Madame et Messieurs

3) Pronouns that occur before the verb but after another pronoun in the phrase, such as le in c) and me in e). These pronouns are the direct or indirect object, and they are coded PS.

4) Pronouns that occur after the verb are either objects or pronouns connected by a hyphen to the verb, such as moi in b), tu in g). These pronouns are coded PT.

This revision in the coding of pronouns has been made in order to generate unambiguously those sentences with more than one personal pronoun. In Philippe I [9], where all personal pronouns were coded PE, there was a significant amount of ambiguity.

En is coded PS and also DV (adverb) and EP (preposition), to account for its various uses.

y (which was coded PE,DV in Philippe I [9]) is coded PY. This allows the grammar to identify the common construction il y a or y a.

As already said, the other four categories of pronouns have the same coding as in Philippe I [9]. The entire coding system is given in Table I.

II.4 Adjectives

The six main subcategories of adjectives are: qualitative (AQ), numerical (AC for cardinal, AN for ordinal), demonstrative (AD), interrogative (AT), possessive (AO), indefinite (AI).

II.4.1 Qualitative adjectives

Qualitative adjectives are coded according to gender, number and position (1 for pre-position, 2 for post-position, 0 for undetermined). In the case of gender and number the three-fold classification illustrated for coding common nouns has been used. For instance, blanche is coded AQ2.0.2, rouges AQ0.0.2.

In Philippe I [9] most past participles were coded as qualitative adjectives and verbs (AQ,V). In the full grammar, this resulted in spurious ambiguities since sentences involving past participles were always ambiguous. For example, the coding of elle est cassée was PF VC,VA AQ,VN. The grammar recognized PF VC AQ as well as PF VA VN. At the final step, after statistical disambiguation, the string PF VA VN was eliminated.

This solution was correct from both a syntactic and semantic standpoint when the sentence was simple. However, it was unsatisfactory for sentences with prepositional phrases, such as for example elle est tombée dans l'eau, where the probabilistic disambiguation generated the type PF VC AQ EP DN NC. Hence, most past participles are coded as verbs.

II.4.2 Numerical adjectives

Cardinal adjectives are coded AC.
Ordinal adjectives are coded AN; their gender and number is taken into account. For example, premier is coded AN1.0.

The coding for demonstrative, interrogative and possessive adjectives is given in Table I.

II.4.3 Indefinite adjectives

In our first work, indefinite adjectives were coded AI,AQ. Since after probabilistic disambiguation, AQ was always chosen instead of AI, we decided to code them as AI. Morever, indefinite adjectives which cannot be differentiated from indefinite pronouns when there is no reference to the context are coded PI,AI. This is the case, for instance, for autre, même, quelque. Furthermore, tout and toute which can modify an adjective, an adverb or a verb are also coded DM.

II.5 Verbs

In written French most verbs have a different ending for each

form. These variations of inflectional patterns are not reflected in the spoken language. For example (j') aime, (tu) aimes, (ils) aiment do not differ phonetically. Our coding (presented in Table II) which agrees to a large extent with Martinet [4] is based on the phonetic pattern of verb endings.

The main distinction is among transitive (VT), intransitive (VI), auxiliary (VA), impersonal (VM), pronominal (VR), semi-auxiliary (VD) verbs, and verbs which may be either intransitive or transitive (VN).

Getting the coding to reflect the mode, tense, and person of the inflectional patterns of spoken French was the most difficult aspect of our task. The examples given in Table III show the variety of these inflections.

II.5.1 Infinitive

The infinitive form is coded 0, indicating its indeterminate status. Imperative (second person plural), past participle or second person plural of the present indicative are coded also 0 when they do not differ phonetically from the infinitive. For example arrêter, arrêtez and arrêté are all coded VN0. (1)

(1) The examples given for the coding of verbs are not necessarily extracted from the corpus, since several forms were not used by Philippe.

II.5.2 Present indicative

The three singular forms as well as the third person plural are coded 1 when they do not differ phonetically from each other. For example the forms (je) mange, (tu) manges, il(s) mange(nt) are all coded VT1. When the second person and the third person singular differ from the first person singular, they are coded 2. For example, (j)'ai is coded VT1,VA1, (tu) as and (il) a are coded VT2,VA2. The first person plural (ending ons) is coded 3, as in (nous) voyons VN3. The second person plural (ending ez) is coded 4 when it differs phonetically from the infinitive; for example (vous) finissez is coded VN4, while (vous) chantez is coded VN0 (same phonetic ending as infinitive). When the third person plural differs phonetically from the three singular personal forms it is coded 5; for example (ils) vont is coded VI5,VD5, (ils) servent is coded VT5, while (ils) chantent is coded VN1.

II.5.3 Imperfect

Concerning the imperfect ("imparfait de l'indicatif"), the three singular forms and the third person plural do not differ phonetically from each other. They are all coded 6. For example, (je) fermais, (tu) croyais, (il) tenait and (ils) montaient are coded VN6. The first and second person plural (ending with ions and iez, respectively) are coded 7 and 8 when they differ phonetically from the first and second person plural of the present indicative (or from the infinitive for the second person plural). For example, (nous) trouvions is coded VI7, (vous) arriviez is coded VI8, while (nous) riions should be coded VI3 and (vous vous) méfiiez should be coded VI0.

II.5.4 Future Indicative

For the future indicative, the first person singular (ending rai) and the second person plural (ending rez) are coded 9: for example, (j') attraperai and (vous) arriverez are coded VI9 and VI9, respectively. The second and the third person singular (endings ras and ra) are coded 10; (tu) partiras is coded VI10, (il) sortira VN10. The first and third person plural (endings rons and ront) are both coded 11; for example (nous) danserons is coded VN11, (ils) démoliront VT11.

II.5.5 Present participle

The present participle is coded 12; for example (en) enlevant is coded VT12.

II.5.6 Conditional

The three singular forms (endings rais, rais, rait) and the third person plural (ending raient) of the conditional present ("conditionnel present") were difficult to distinguish, in Philippe's speech, from the first person singular of the future. Consequently, they are all coded 9. For example (il) soignerait and (ils) mangeraient are coded VT9. The first and second person plural (endings rions and riez) are coded 13 and 14, respectively; for example (nous) recevrions, VT13, and (vous) prendriez VN14.

II.5.7 Simple past

The simple past ("passé simple") is rarely used in spoken language and, as we anticipated, was not used by Philippe, at least not in the first and the second person plural. The three singular forms are coded 18 when they differ phonetically from the

other verbal forms which have already been assigned a lower number; for example (tu) mangeas is coded VT18, (il) prit VN18. Verbs for which the first person singular of the simple past ends with ai are difficult to differentiate from the first person singular of the imperfect (ex: je mangeais, je mangeai). Consequently, they receive the same code as the first person singular of that tense, namely 6. Verbs for which the three singular forms of the simple past are the same as the first person present indicative are coded 1; for example, (tu) finis is coded VN1. Finally, the third person plural of the simple past is coded 19; for example (ils) dormirent is coded VI19, (ils) reçurent, VT19.

II.5.8 Present subjunctive

We turn now to the present subjunctive ("subjonctif présent"). For many verbs, the singular forms of the present subjunctive are the same phonetically as for the present indicative. When the three singular forms and the third person plural of the subjunctive are similar phonetically to the first person singular of the present indicative, they are coded 1. When they are the same phonetically as the third person plural of the present indicative they are coded 5. For example (que je) chante is coded VN1, (qu'il) finisse is coded VN5. Verbs with a specific form for the first, second and third person singular are coded 15; for example, (que j') aille, (que tu) ailles, (qu'il) aille are coded VI15. The first and second person plural of the present subjunctive are in many cases the same as the first and second person plural of the imperfect indicative. Consequently, they are coded 7 and 8 respectively. For example (que nous) réussissions is coded VN7, and (que vous) réussissiez, VT8. However, when the first and second person plural differ from the imperfect, they are coded 16 and 17; for example (que nous) ayons is coded VT16,VA16, (que vous) ayez, VA17,VT17.

II.5.9 Past participle

When the past participle is not different from the infinitive, it is coded 0. For example, fermé is coded VN0. When the past participle is the same phonetically as the first person singular of the present indicative, it is coded 1. For example, fini is coded VN1, dit VT1. When the past participle is different from both infinitive and present indicative, it is coded 20, even when it does not differ phonetically from the simple past. For instance, reçu and partis are coded VT20 and VI20, respectively. In this one instance, we did not follow the basic rule of always coding with the smallest number, since the simple past is less frequently used by Philippe than the past participle.

II.5.10 Imperative

Finally, we turn to the imperative. The imperative forms are the same phonetically as the present indicative; and hence they were coded the same when no reference was made to the context. However, note that in written language a hyphen is used to link the verb to a personal pronoun object, as in donne-moi. As we mentioned earlier, the imperative-personal pronoun is identified and treated as a unit. In these cases, the imperative form is coded 21, followed by the appropriate code for the personal pronoun, which is usually PT. For example, donne-moi is coded VN21-PT.

Several aspects of the coding of verbs need to be emphasized. In the first place, we have coded with the smallest number phonetically possible for a given form. Thus, the past participle allé or the imperative allez are coded 0. The exception to this rule is in those

verbs for which the past participle has the same phonetic form as the simple past. Example: je pris, pris. The other point is that the individual words of the vocabulary are coded independently of their context and, consequently, composed forms, such as j'aurai reçu are not coded as a single verb form but rather the individual words aurai and reçu are coded separately.

Finally, we did not devise a coding for some tenses that are extremely infrequent in spoken speech, such as the imperfect subjunctive ("imparfait du subjonctif").

II.6 Adverbs

In Philippe I [9], we distinguished 7 subcategories of adverbs: indetermined (DV0), adverbs of time (DV1), adverbs of location (DV2), adverbs of quantity (DV3), adverbs of quality ("de manière") (DV4), adverbs of affirmation (DV5), adverbs of negation (DV6), and adverbs of interrogation (DV7).

In the analysis of the full grammar, described in Section III, the subscripts were removed, and all adverbs functioned identically. Since it was indispensable for the grammatical analysis to identify adverbs of negation, we have coded as DG the negative words pas, plus. Ne, which is a part of the whole negation ne...pas, is coded NE, in order to characterize the few cases in which Philippe used the full form of the negation.

We have introduced a new subcategory of adverbs, DM, which

includes adverbs that can modify an adjective, an adverb or a verb. This is the case for assez, trop, peu, which are coded DV,DM. Très is coded DM only, because it always modifies an adjective, except when it is an answer to a question, as in Tu l'aimes? Oui, très.

II.7 Prepositions

Prepositions are in the category EP. Ordinary prepositions, such as sur, avec, are coded EP0. The special cases of the forms of de and à are coded EP1 through EP7 as shown in Table I.

II.8 Conjunctions of subordination

All conjunctions of subordination, such as puisque, parce que, are coded CS without any elaboration of subcategories.

II.9 Conjunctions of coordination

Conjunctions of coordination are coded CC. Philippe used et; in a few cases, he used ou and ni.

II.10 Interjections and onomatopeias

In the broad category of interjections and onomatopeias are classified such words as ah, oh, eh, ben... which are coded KO. Many of these words are onomatopeias that children frequently use when they are playing.

II.11 Locutions

There are four subcategories: indetermined (LC0), adverbial locution (LC1), prepositive locution (LC2), interrogative locution (LC3). Once the subscripts are eliminated, this category of locutions is rather heterogeneous, and any detailed analysis cannot be based on the categories given here. For example, courtesy words (merci, pardon, bonjour...) are within the same category as est-ce (part of the interrogative expression qu'est-ce que...), là-dessus, là-dedans.

II.12 Conclusion

A given word may function in different ways, depending upon its context. Words of this kind have been assigned to two or more categories. This is denoted by commas in the coding. For instance cours may be the plural form of a substantive, or the first person singular of the verb courir. Consequently it is coded NC2.0,VI1. When this occurs, we say that the word is multiply classified. We have not included either all or only those multiple classifications that are in all likelihood present in Philippe's speech but we have tended to be somewhat conservative in our use of multiple classification. When it was highly unlikely that Philippe would have used a certain word in more than one context, we did not multiply classify that word. For example, the word mets has been coded as a verb but not as a noun, since it is our belief that the latter usage ("something related with

food") was unlikely to occur. Furthermore, when we were in doubt about word usage, we used several computer programs to check whether or not Philippe used a word in more than one sense. Thus words which had previously been multiply classified were assigned only one code. On the basis of observed usage, for instance, we checked that rentrée was always used as a verb, never as a common noun, and consequently it is coded as VN. This procedure enabled us to decrease the lexical ambiguity that would further complicate the construction of a grammar.

TABLE I

CODING OF GRAMMATICAL CATEGORIES

ARTICLES

Indefini	Defini
IN 1 un	DN 1 le, l'
IN 2 une	DN 2 la, l'
	DN 3 les

PRONOUNS

PERSONAL

PF	je, j'	PF,PS,PT	lui
PS	me, m'	PF,PT	eux
PF,PT	moi	PS	se, s'
PF,PS,PT	nous	PT	soi
PF	tu	DN,PS	le, la, les, l'
PS	te	PS,PO,AO	leur
PF,PS	t'	PS,DV,EP	en
PF,PT	toi	PY	y
PF,PS,PT	vous		
PF	il, ils		
PF,PT	elle, elles		

INTERROGATIVE OR RELATIVE

PR 1	qui	PR 7	auquel, auxquels, auxquelles
PR 2	que	PR 8	duquel
PR 3	quoi	PR 9	lesquels, lesquelles
PR 4	qu'	PR 10	desquels, desquelles
PR 5	lequel	PR 11, DV	où
PR 6	laquelle	PR 12	dont (relatif)

DEMONSTRATIVE

PD 1	celui	PD 8	ceux-ci
PD 2	celle, celles	PD 9	ceux-là
PD 3	ceux	PD 10	ceci
PD 4	celui-ci	PD 11	cela
PD 5	celui-là	PD 12	ça

PD 6 celle-ci, celles-ci
PD 7 celle-là, celles-là
PD 13 c'
PD 14 ce

INDEFINITE

PI 1 on
PI 2 quelqu'(un)
PI 3 quelques-uns
PI 4 quelques-unes
PI 5 personne
PI 6 rien
PI 7 chacun
PI 8 autre, etc...

POSSESSIVE

PO 1 (le,les) mien(s)
PO 2 (le,les) tien(s)
PO 3 (le,les) sien(s)
PO 4 (le,les) nôtre(s)
PO 5 (le,les) vôtre(s)
PO 6 (la,les) mienne(s)
PO 7 (la,les) tienne(s)
PO 9 (le,la,les) leur(s)

ADJECTIVES

AQ : qualitative
AC : cardinal
AN : ordinal
AI : indefinite

POSSESSIVE

AO 1 mon
AO 2 ton
AO 3 son
AO 4 ma
AO 5 ta
AO 6 sa
AO 7 mes
AO 8 tes
AO 9 ses
AO 10 nos
AO 11 vos
AO 12 notre
AO 13 votre
AO 14 leur(s)

DEMONSTRATIVE

AD 1 ce
AD 2 cet
AD 3 cette
AD 4 ces

INTERROGATIVE

AT 1 quel(s)
AT 2 quelle(s)

ADVERBS

DV 0 indetermined
DV 1 time
DV 2 location
DV 3 quantity

PREPOSITIONS

EP 0 indetermined
EP 1 de,d'
EP 2 du, d'
EP 3 des, d'

DV 4 quality
DV 5 affirmation
DV 7 interrogation
DG negation
DM modifiers

EP 4 à
EP 5 au,aux
EP 6 de, d' (with verbs)
EP 7 à (with verbs)

CONJUNCTIONS

CS : subordination

CC : coordination

LOCUTIONS

LC 0 indetermined
LC 1 adverbial
LC 2 prepositive
LC 3 interrogative

NOUNS

NC : common noun

NP : proper noun

KO : interjections and onomatopeias

MS : uncodable expressions (without meaning)

TABLE II

Verb Coding Scheme

Basic tense/Mood	Coding	Person/Number
Infinitive and similar phonetic forms	0	
Present indicative	1,2	basic singular forms
	3	1st person plural
	4	2nd person plural
	5	3rd person plural
Imperfect	6	persons sing. and 3rd plural
	7	1st person plural
	8	2nd person plural
Future	9	1st person singular
	10	2nd and 3rd persons sing.
	11	1st and 3rd persons plural
Present participle	12	
Conditional	13	1st person plural
	14	2nd person plural
Subjunctive	15	three persons singular
	16	1st person plural
	17	2nd person plural
Simple past	18	three persons singular
	19	3rd person plural
Past participle	20	
Imperative	21	(when followed by a pronoun)

TABLE III

Examples of Coding for Verbs

Tense/Mood	Coding	Sample French Verbs				
Infinitive	0	arriver	finir	pouvoir	aller	être
Present indicative	1	arrive	finis	peux	vais	suis
	2				vas	es
	3	arrivons	finissons	pouvons	allons	sommes
	4		finissez	pouvez		êtes
	5		finissent	peuvent	vont	sont
Imperfect	6	arrivais	finissais	pouvais	allais	étais
	7	arrivions	finissions	pouvions	allions	étions
	8	arriviez	finissiez	pouviez	alliez	étiez
Future	9	arriverai	finirai	pourrai	irai	serai
	10	arriveras	finiras	pourras	iras	seras
	11	arriverons	finirons	pourrons	irons	serons
Present participle	12	arrivant	finissant	pouvant	allant	étant
Condi-tional	13	arriverions	finirions	pourrions	irions	serions
	14	arriveriez	finiriez	pourriez	iriez	seriez
Subjunctive	15				aille	sois
	16					soyons
	17					soyez
Simple past	18					fus
	19	arrivèrent	finirent	purent	allèrent	furent
Past participle	20			pu		été
Imperative	21		finis-le			

Chapter III

GRAMMAR

III.1 Group 1 Highest Level Rules

S is the starting symbol of the rewrite rules of the grammar.

Three main kinds of utterances are generated from S:

> a) utterances of length 1 or 2 consisting of adverbs, locutions, interjections and numerals; b) noun phrases and adjective phrases which stand alone, i.e. without a verb; c) utterances in which a verb is present, and questions (with or without a verb).

Rule 1.1 S -> S1 parses 16 types (2611 tokens)

Out of the tokens, 2199 are adverbs. This rule is rewritten in group 2.

Rule 1.2 S -> S2 330 types (2669 tokens)

This rule recognizes utterances of variable length, from one to seven words, which do not contain a verb.

The following types of one-word utterances are recognized:

246	nc (common noun)	caca
137	np (proper noun)	Papa
97	pd (demonstrative pronoun)	ça
42	aq (adjective)	noir
13	in pi (indefinite pronoun)	un autre
10	pf (possessive pronoun)	moi
6	pi (indefinite pronoun)	rien
6	dn pi (indefinite pronoun)	l'autre
4	dn po (possessive pronoun)	le mien

The longest utterances have, of course, a low frequency.

```
dg dg ep nc ep dn nc
       non pas de volant dans les avions
in aq nc ep ao aq nc
       un petit jouet pour mon petit mouton
```

One utterance of eight words is recognized, but its grammatical analysis is incorrect:

```
dg dg ep dn nc ep dn nc
       non pas dans le préau sous la classe
```

It is interesting to note that common nouns are more frequently used with a determiner than without:

```
295 in nc       une moto
253 dn nc       la voiture
```

```
Rule 1.3        S -> S3         2729 types (6014 tokens)
```

This rule recognizes both utterances that contain a verb and questions. Questions can be restricted to a single word (PR: quoi?) or can have a more complex form. The production rules for the questions are presented in group 28.

The length of the utterances parsed by rule 1.3 varies from one to 15 words:

```
109 vt          manger
 73 pr          quoi
 48 vn          viens
 18 vi          tombé
```

```
dn nc pf vt dn nc ep in nc pr vc ep in nc dv
       le monsieur il cire les chaussures d'un monsieur qui est
       sur un banc dehors

dv pd vc in nc pr vt ep dn nc ep ao nc dv
       là c'est un camion qu'envoie de l'eau avec son balai dessous

pf pf vd vt ep np cc pf vd ps vn ep dn nc
       moi je vais emporter chez mamie et je vais le donner
       à les poules.
```

III.2 Group 2 Incomplete Utterances

Most of the rules in this group generate terminals directly, the exceptions are rules 2.2 (S1 -> DV') and 2.6 (S1 -> NUM). The rules have only one or two branches; they parse the following grammatical categories introduced in the dictionary: adverbs (DV,DG), locutions (LC), onomatopeias and interjections (KO) and several combinations of these categories (KO DG, DG DV, DG DG, DG LC, CC DG, EP DV). Rule 2.15 (S1 -> DM DV) is intended to parse an adverb modified by one another, as in tout bêtement. This rule was not used because in the course of the disambiguation the indefinite pronoun PI (which is one of the coding of tout) was chosen over DM when a word, such as tout had PI,DM as its multiple classification. Rule 2.16 (S1 -> EP DV) recognizes utterances formed by a preposition and an adverb, such as par là. Rule 2.17 (S -> PS DV) recognizes utterances in which a personal pronoun is followed by an adverb, as in le voilà.

Rules 2.2 (S1 -> DV') and 2.6 (S1 -> NUM) use non-terminal symbols and have two nodes. Examples are un peu, un petit peu and deux.

III.3 Group 3 Noun Phrase and Adjective Phrase Utterances

This group recognizes noun phrases and adjective phrases which are not combined with a verb. We had two motivations for generating with separate rules noun phrases that occur as the complete sentence

without a verb: first, we wanted to improve the probabilistic fit of the grammar; second, we wanted to reveal a developmental trend in the usage of noun phrases.

Rule 3.1 S2 -> SN*1 150 types (2025 tokens)

This rule is the basic rewrite rule of the noun phrases. We have already given some examples when presenting rule 1.2. Other examples are:

"le film des Aristochats"
"la chambre de Misette"
"avec une allumette"

Rule 3.2 S2 -> DV' SV 1 type (1 token)

This rule recognizes noun phrases beginning with un peu de or un petit peu de. Both expressions function as adverbs of quantity modifying a noun phrase.

"un petit peu de l'eau"

Rule 3.3 S2 -> SN DV 25 types (100 tokens)

This rule recognizes noun phrases ending with an adverb:

"dans les tuyaux là"

Rule 3.4 S2 -> SN NP 16 types (54 tokens)

This rule recognizes noun phrases ending with a vocative:

"du chocolat papa"

This rule recognizes also noun phrases such as:

"la tour Montparnasse"

We believe that compound proper nouns should not be handled as separate words by the grammar, but should be instead a function of a

prescanner that corrects the orthography. For details of such a scanner, used for a question-answering system, see [6]. Since utterances that contain a proper noun following a common noun, as in la tour Montparnasse, are not very frequent in the corpus, and are not of great interest from either the grammatical or semantic point of view, we did not find it necessary to utilize the prescanner technique. For the same reasons, we did not adopt the solution which consists of adding new rules linking together the common noun and the proper noun.

Rule 3.5 S2 -> DV SN*1 37 types (162 tokens)

This rule recognizes noun phrases beginning with an adverb:

"dedans le camion"
"encore du chocolat"

Rule 3.6 S2 -> DG SN*1 23 types (63 tokens)

Examples:

"pas les jambes"
"non un gros bateau"
"non du chocolat"

Rule 3.7 S2 -> KO SN*1 17 types (41 tokens)

This rule recognizes noun phrases beginning with an interjection:

"oh le trou"
"ben dans une salle"
"ben à la gare"

Rule 3.8 S2 -> LC SN*1 6 types (13 tokens)

This rule recognizes noun phrases beginning with a locution; the locution is either adverbial, or a formula of courtesy:

"là-haut médicaments"
"bonjour petits amis"
"merci gentil papa"

Rule 3.9 S2 -> CC SN*1 14 types (52 tokens)

This rule parses noun phrases beginning with a conjunction of coordination:

"et ça"
"et la dame"
"et des tiroirs"

Rule 3.10 S2 -> DG DG SN*1 9 types (11 tokens)

In the noun phrases recognized by this rule, the first adverb of negation is non and the second one is pas:

"non pas le ventre"
"non pas samedi"
"non pas avec de la colle"
"non pas avec Ginette"

Rule 3.11 S2 -> CC DV SN*1 4 types (9 tokens)

Examples:

"et là ça"
"et comme ça"
"ou comme ça"
"et puis une grenouille"

Rule 3.12 S2 -> CC DG SN*1 3 types (4 tokens)

Examples:

"et pas deux yeux"
"et pas deux petits trous"
"mais non dans l'autre sens"

Rule 3.13 S2 -> DG SN*1 DV 4 types (9 tokens)

Examples:

"plus rien dedans"
"plus musique là"
"non au milieu là"
"non un gros camion dedans"

Rule 3.14 S2 -> CC SN*1 DV 6 types (9 tokens)

Examples:

"et la remorque là"
"et des groseilles aussi"

Rule 3.15 S2 -> SA 4 types (62 tokens)

Adjectives are recognized by this rule:

"noir"
"rouge"
"gros"

Rule 3.16 S2 -> DET SA 4 types (23 tokens)

Adjectives preceded by a determiner are recognized by this rule:

"une grande"
"une petite"

Rule 3.17 S2 -> DET SA CC DET SA 2 types (10 tokens)

"la petite et la grosse"

Rule 3.18 S2 -> DG SA 2 types (9 tokens)

"pas vrai"
"pas chaud"
"pas très belle"

Rule 3.19 S2 -> CC SA 1 type (2 tokens)

"et noire"

Rule 3.20 S2 -> DV SA 2 types (10 tokens)

"encore rouge"
"beaucoup glacée"

III.4 Group 4 Utterance Combination

Rule 4.1 S3 -> S' 1808 types (4385 tokens)

This rule recognizes utterances which have a verb whether or not combined with other words.

The symbol S' is a non-terminal of high level which is rewritten as a verb phrase, as a complete sentence, or as several sentences which are linked by a relative pronoun or a conjunction of subordination or of coordination.

Rule 4.2 S3 -> DG S' 120 types (176 tokens)

This rule introduces an adverb of negation at the beginning of the sentence:

"non je veux pas"
"pas chercher les voitures"
"non elle est rouge"

Rule 4.3 S3 -> DV S' 327 types (571 tokens)

Most of the utterances recognized by this rule begin with a question word:

"où il est le fil?"
"où ils sont?"
"comment on les fabrique les bols?"
"pourquoi c'est dangereux?"

In some cases the adverb is not a question word, as in:

"dedans il est blanc"
"oui il est drôle"

Rule 4.4 S3 -> LC S' 8 types (10 tokens)

This rule parses sentences which begin with the conjunction of subordination parce que, where the que has been omitted. Such conjunctions were coded with a specific subscript; since the subscripts are not taken into account by the grammar, sentences which begin with an adverbial locution are also recognized:

"parce il lui tombe de la farine"
"parce j'en ai besoin des lumières"
"parce ils sont morts"
"parce ils sont gentils les oiseaux"

"là-bas j'ai une place"
"là-dedans a un gros camion"

Rule 4.5 S3 -> CC S' 26 types (33 tokens)

"et ça c'est la maison"
"et ça ça roule"
"et ça ça baisse et ça ça revient"
"et le livre qui tombe tout seul"

Rule 4.6 S3 -> KO S' 112 types (146 tokens)

"ah c'est chaud"
"oh regarde"
"ah ton stylo est tombé"

Rule 4.7 S3 -> CS S' 3 types (15 tokens)

"s'il te plaît"
"puisque c'est prêt"

Rule 4.8 S3 -> S' KO 13 types (16 tokens)

"c'est long hein"
"ça fait ouah"
"ça fait croq"

Rule 4.9 S3 -> LC CS S' 40 types (55 tokens)

This rule is equivalent to rule 4.4, the difference being that the complete compound form parce que is recognized:

"parce qu'elle est sale"
"parce qu'y aura pas la lumière"

Rule 4.10 S3 -> CC DV S' 51 types (60 tokens)

"mais où il est?"
"et là il téléphone"
"mais si tu as fait chauffer"
"et puis il s'est noyé là"

Rule 4.11 S3 -> DG DG S' 11 types (11 tokens)

"non pas ta culotte l'as pas raccourcie"
"non non c'est pas ça"
"non non je veux pas le boire"
"non non je veux que tu restes un peu encore"

Rule 4.12 S3 -> KO DV S' 15 types (20 tokens)

"oh comme c'est rigolo"
"ah comment on fait?"
"ah encore tombé"
"oh oui a un trou"

Rule 4.13 S3 -> CC DG S' 12 types (12 tokens)

"mais non c'est pas le chausson"
"mais non c'est pour conduire"
"mais non ça suffit pas"

Rule 4.14 S3 -> DV DV S' 14 types (16 tokens)

This rule and the following have been introduced to parse sentences beginning or ending with aujourd'hui. Since the two graphic elements of that adverb have been coded separately, aujourd is coded DV and hui is coded DV; consequently the coding of aujourd'hui is DV DV.

"aujourd'hui je vais à l'école"
"aujourd'hui je crois pas qu'elle vient Ginette"

In addition, these two rules parse utterances in which two consecutive adverbs appear:

"là aussi c'est chez Mamie"
"alors là ils sont partis là"

Rule 4.15 S3 -> S' DV DV 25 types (34 tokens)

"je mange tout seul aujourd'hui"
"ginette elle dit encore ouais"
"c'est où là?"

Rule 4.16 S3 -> CC DV DV S' 4 types (4 tokens)

"et puis alors un chasseur qui est venu"
"et là aussi c'est aussi des toboggans"
"et puis après elle tombe la voiture"

Rule 4.17 S3 -> Q 141 types (458 tokens)

Rule 4.17 generates questions. It should be mentioned that not

all the possible types of questions are analyzed by this rule. We have already seen that questions beginning with an adverb were parsed by rule 1.3. Furthermore, indirect interrogative sentences and questions which have the same superficial structure as affirmative sentences are analyzed by rule 4.1. Thus, interrogative forms differing from declarative forms only by the presence of a question mark are not recognized as questions. (1) Moreover, since subscripts of the dictionary are ignored by the grammar, we do not differentiate adverbs of interrogation from other adverbs. Thus,

"où il est mon stylo à moi?"
"pourquoi t'as laissé tes chaussures dehors?"
"maintenant je vais boire avec la cuiller"
"oui je crois qu'on peut ouvrir"

are parsed by rule 4.1.

III.5 Group 5 Utterances with a Verb

This group presents the rewrite rules for utterances which have a verb.

Rule 5.1 S' -> SV 601 types (1631 tokens)

This rule recognizes complete sentences, and also incomplete sentences, in which the subject is missing:

"regarde"
"mettre dans le bol le sucre"
"regarde toutes les fleurs là"
"a des sous dans mon porte-monnaie"

Rule 5.2 S' -> SNS SV 1760 types (3672 tokens)

(1) Sentences that were deemed questions were terminated with a question mark during transcription. Our analysis does not take into account the punctuation of the sentences.

This rule is the basic rule for sentences which have a subject:

"j'enlève tout le fil"
"je sais pas"
"j'ai des petites boules"

Rule 5.3 S' -> SNS SV CS SNS SV 74 types (82 tokens)

This rule recognizes sentences in which a subordinate clause is linked to the main clause, such as:

"je le cache puisque personne le sait"
"je veux que tu restes encore un peu"

Rule 5.4 S' -> SNS SV CC SNS SV 13 types (15 tokens)

This rule recognizes independent sentences linked by coordinative conjunction:

"et ça ça baisse et ça ça revient"
"et puis on fait le pain et on met de la croûte autour"

Rule 5.5 S' -> SNS SV PR SV 73 types (81 tokens)

This rule recognizes sentences in which the object of the first verb is the antecedent of the relative pronoun subject of the subordinate clause:

"et ça c'est un camion qui roule dans la neige"
"ça c'est le chien qui était dans les Aristochats"

Rule 5.6 S' -> SNS PR SV 31 types (35 tokens)

In the sentences recognized by this rule, the relative clause modifies the noun phrase (SNS):

"quelque chose qui est un peu noir"
"et le livre qui tombe tout seul"

Rule 5.7 S' -> SN** SNS SV 59 types (66 tokens)

This rule recognizes sentences beginning with a prepositional phrase (which is ordinarily put at the end of the sentence):

"dans le sac je le mets"
"avec un couteau je vais te la couper"
"en Angleterre elle est"

It also recognizes sentences in which the direct or indirect object occurs at the beginning:

"des cacahuètes ils mangeaient"
"à quoi ça sert?"

Rule 5.8 S' -> SN** GV' 10 types (14 tokens)

This rule recognizes sentences in which the verbal group modifies the noun phrase:

"de l'eau à boire"
"des boutons pour conduire"

and sentences in which there is an inversion of the prepositional phrase:

"sur une échelle pour le faire"

Also this rule recognizes two sentence types with frequency 1 that begin with en train de... followed by an action verb.

Before presenting the rewrite rules of groups 6, 7 and 8, we shall explain the reason for introducing these three non-terminal symbols: SN, SNS and SNP, which have similar rewrite rules. These non-terminals have been introduced to differentiate the three main functions of the noun phrase: subject, object, nominative predicate. We wanted to formulate correctly the semantics associated with each rule, and also see if there was a developmental trend in the usage of noun phrases depending on their function. In the first grammars we

wrote, the noun phrase was derived using only one rule: S -> SN, and it was impossible to separate the noun phrases by their function. Furthermore, the same derivations were obtained for a noun-phrase object of a transitive verb, as in il fume une cigarette, as for a noun-phrase subject of an intransitive, when the subject was repeated at the end of the sentence, as in il marche l'hélicoptère. In order to have different derivations depending upon the nature of the verb, we introduced the non-terminals SN, SNS and SNP.

SN parses noun phrases which are nominative predicates of the copula verb. SNS parses noun phrases which are the subject of a verb, whatever its nature, SNP parses noun phrases which are the object of the verb.

We have not been entirely successful in our attempt to analyze the three main functions of the noun phrase using specific rules. Instances of failure are when the subject is analyzed by a rewrite rule of SNP (ex: fume la maîtresse), and also when the object is analyzed by a rewrite rule of SNS (ex: un sucre je veux).

By adding two more rules to the grammar, such sentences would correctly be analyzed.

We would need SNP -> SNP PF for parsing un sucre je veux and SV -> V SNS for parsing fume la maîtresse. This would, however, result in two derivations for each of these two types.

III.6 Group 6 Nominative Predicate Noun Phrases

SN is rewritten as SN*, SN**, PD, PT.

Rule 6.1 SN -> SN* 372 types (892 tokens)

When rewritten as SN*, SN parses noun phrases (with or without a determiner) which are nominative predicates of the copula:

"c'est une petite fille"
"c'est Papa"

or objects of an existential verb:

"c'est samedi".

SN is also used to rewrite rules 3.3 and 3.4.

Rule 3.3 S2 -> SN DV

"une voiture là"

Rule 3.4 S2 -> SN NP

"une cigarette Madeleine"

In these two cases, a verb has been omitted; we do not speculate about the nature of the verb which may be implicit in the semantics.

Rule 6.2 SN -> SN** 150 types (246 tokens)

When rewritten as SN** , SN parses noun phrases beginning with the preposition EP which occurs in sentences containing a linking verb. Recall that EP can be a preposition in the classical sense, or an article such as de, du, des, d'. Consequently, SN** will be either a nominative predicate or a prepositional noun phrase.

"c'est des jouets"
"c'est dans le tiroir"
"c'est pour papa"
"c'est chez Mamie".

SN introduces PT and PD when they are the nominative predicate.

Rule 6.3 SN -> PT 16 types (31 tokens)

"c'est moi"

Rule 6.4 SN -> PD . 49 types (148 tokens)

"c'est quoi ça?"
"c'est celui-là"

III.7 Group 7 Subject Noun Phrases

Rule 7.1 SNS -> SN* 521 types (789 tokens)

This rule parses the subject of the verb:

"on casse"

or a repetition of the subject:

"où elle va Maman ?"

When SN* is rewritten as NP, it is often a vocative:

"enlever le toboggan Maman"

but in a few cases NP is the subject:

"brûlé l'assiette Mamie"

Rule 7.2 SNS -> PF 1011 types (2235 tokens)

PF is the coding for personal pronouns which are the subject of the verb.

Rule 7.3 SNS -> SN* PF 138 types (161 tokens)

This rule parses sentences in which SN* is an apposition to the personal pronoun subject of the verb:

"Philippe il rentre"
"le plancher de Janou il tient"

However, in some cases, SN* is the direct object of the verb:

"un jouet je veux"
"un sucre tu manges"

Rule 7.4 SNS -> PF PF 46 types (55 tokens)

In most cases, the first PF is in apposition to the second one and emphazises it:

"moi je la connais"
"moi je te chercherai"

In some cases, the second PF ("lui" in the following example) corresponds to an indirect object, and it is a failure of our grammar to use rule 7.4, in such cases as:

"parce il lui tombe de la neige"

where "lui" (coded PF,PS,PT) is analyzed as PF.

Rule 7.5 SNS -> PD 424 types (909 tokens)

PD is the non-terminal symbol for demonstrative pronouns and this rule recognizes PD as the subject:

"ça pique aussi".

In some cases, with intransitive verbs, it appears at the end of the sentence and is the subject:

"marche pas celle-là"

Rule 7.6 SNS -> PD PD 84 types (138 tokens)

The first PD is in apposition to the second and emphasizes it:

"ça c'est un écureuil"
"celui-là c'est le bras gauche".

Rule 7.7 SNS -> SN* PD 22 types (22 tokens)

SN* PD recognizes sentences of the following form:

"cette machine-là ça dit pas quelque chose"
"cette dame-là c'est Madeleine"

SN* is an apposition to PD.

Rule 7.8 SNS -> PD DV PD 5 types (6 tokens)

"ça aussi c'est chez Mamie"

In the 5 types which are recognized the terminal word for DV is aussi.

Rule 7.9 SNS -> SN* DV PF 12 types (12 tokens)

"Maman aussi elle coud"

In some cases, DV is an interrogative:

"Papa où tu vas ?"
"et Maman comment elle s'appelle ?"

In few other cases, SN* is the object:

"Jean-Michel aussi je l'aime bien"
"ton stylo là, j'en ai besoin".

III.8 Group 8 Object Noun Phrases

The main difference between SNS (group 7) and SNP (group 8) is that the latter parses prepositional noun phrases while the first does not. A second difference is that SNP is used for parsing noun phrases that occur with verbs other than the copula.

Rule 8.1 SNP -> SN* 845 types (1593 tokens).

In most cases, this rule parses noun phrases which are the direct object of the verb. However, as we have already seen, in some cases it is the subject:

"font les dames"
"fume la maîtresse"

When SN* is rewritten as NP, it is often ambiguous and its grammatical function cannot be analyzed without reference to the context:

"attends Papa"

Papa can be vocative, if Philippe is speaking to his father, or an object if Philippe is asking somebody else to wait for his father.

Rule 8.2 SNP -> SN** 413 types (708 tokens)

SN** parses noun phrases beginning with EP, which are either the direct object:

"manger des champignons"

or a prepositional noun phrase:

"il monte à l'arbre"
"il roule par la clef".

Rule 8.3 SNP -> PR 15 types (18 tokens).

PR is always a direct object, and the terminal word is quoi.

"pour faire quoi ?"
"je sais pas quoi"

Rule 8.4 SNP -> PD 85 types (173 tokens)

In most cases, PD is a direct object:

"tailler celui-là Maman"

In some cases, it is a repetition of the subject:

"ça parle pas ça"
"ça tourne aussi ça".

III.9 Group 9 Basic Noun Phrases

As we have seen, SN* is a rewrite rule of SN, SNS and SNP. SN* is rewritten by 7 different rules which are the basic rules for recognizing noun phrases which are a subject, object or nominative predicate. As we have mentioned earlier, SN* is also a rewrite rule for noun phrases that end with an adverb or a vocative.

Rule 9.1 SN* -> NP* 1526 types (2969 tokens)

This rule needs no comment. It is the basic rule for recognizing noun phrases:

"je cherche le crayon bleu"
"il traire sa chèvre"
"j'ai vu Danielle"

Rule 9.2 SN* -> NP* CC NP* 12 types (14 tokens)

In all cases but one, the terminal word for CC is et:

"ils sont pas là la grenouille et l'alligator"
"avait Misette et Gronomme dans ce cinéma"
"c'est ses pattes et ses griffes"
"j'ai deux porte-monnaie et deux montres"
"c'est Papa ou Ginette"

Rule 9.3 SN* -> NP* EP NP*

We do not present the number of types and tokens recognized by this rule, since we believe that information misleading in the present case. Many utterances recognized by this rule are also recognized by using rules 8.1 and 10.1; consequently for many utterances we have two different trees. An example will illustrate the necessity of having two derivations for the same string. For example, the string PF VT DN NC EP NC which corresponds to the actual utterance j'enlève les miettes de croissant could also correspond to j'enlève les miettes du lit. While in the first case EP links together miettes and croissant, in the second case EP links the noun to the verb. We have one tree which corresponds to the first analysis, and a second tree which corresponds to the second analysis. j'enlève les miettes de croissant is correctly

analyzed by Rule 9.3, whereas j'enlève les miettes du lit is correctly analyzed by rules 8.1 and 10.1: (2)

PF VT DN NC EP NC

	S		S
(1.3)	S3	(1.3)	S3
(4.1)	S'	(4.1)	S'
(5.2)	SNS SV	(5.2)	SNS SV
(7.2)	PF SV	(7.2)	PF SV
(18.1)	PF GV	(18.1)	PF GV
(21.35)	PF V SNP SN**	(21.25)	PF V SNP
(25.1)	PF VT SNP SN**	(25.1)	PF VT SNP
(8.1)	PF VT SN* SN**	(8.1)	PF VT SN*
(9.1)	PF VT NP* SN**	(9.3)	PF VT NP* EP NP*
(11.14)	PF VT DET NC SN**	(11.14)	PD VI DET NC EP NP*
(16.4)	PF VT DN NC SN**	(16.4)	PF VT DN NC EP NP*
(10.1	PF VT DN NC EP NP*	(11.15)	PF VT DN NC EP NP"
(11.15)	PF VT DN NC EP NP"	(13.5)	PF VT DN NC EP NC
(13.5)	PF VT DN NC EP NC		

Rule 9.3 recognizes the possessive case:

"fais la main de Maman, Madeleine"

Rule 9.4 SN* -> NP* CC NP* EP NP* 2 types (2 tokens)

"regarde la voiture et les roues de la voiture"
"avait Misette et Gronomme dans ce cinéma"

The two utterances have each two different trees. While rule 9.4 analyzes correctly the first utterance, rules 8.1 and 9.2 will be chosen to analyse the second one.

Rule 9.5 SN* -> NP* EP EP NP* 9 types (10 tokens)

This rule recognizes utterances in which the first EP is pour ou avec, and the second one is des:

"un magasin pour des musiques là"

(2) The two sequences of strings are derivations of the lexical form starting from the start symbol S. The parenthesized numbers show which rule is involved at each step.

"comment c'est fait les confitures avec des myrtilles?"
"ça c'est une voiture avec des ressorts"
"Papa il m'a acheté une voiture avec des roues"

It should be mentioned that this rule also recognizes utterances in which there is a repetition of the preposition:

"c'est le papa de des petits cochons là"
"j'ai un pistolet dans dans le pantalon"

Rule 9.6 SN* -> NP* EP PT 13 types (13 tokens)

This rule is "ad hoc". It recognizes utterances in which Philippe used the expression à toi, à moi, instead of using a possessive adjective, or for reinforcing the idea of possession:

"où il est le stylo à moi?"
"veux la montre de moi"
"où il est ton ciré à toi?"

Rule 9.7 SN* -> NP* EP PR 3 types (5 tokens)

"c'est le nom de quoi ça ?"
"c'est la chambre de qui ?"

As it is shown by the examples, this rule parses questions in which there is a search for information concerning the nominative predicate.

III.10 Group 10 Prepositional Noun Phrases

As we already mentioned, this group of rules recognizes noun phrases beginning with EP which is the terminal symbol for a preposition. However, all the noun phrases recognized by this rule are not prepositional sentences, since some articles (de, du, des, d') have been coded EP.

Rule 10.1 SN** -> EP NP* 673 types (1142 tokens)

"c'est des roues" (direct object)
"ils montaient sur le chameau" (prepositional phrase)
"elle rentre dans la maison" (prepositional phrase)

Rule 10.2 SN** -> EP NP* EP NP*

This rule corresponds to rule 9.3, with the difference that it begins with EP. Here again we do not present the number of types recognized, since half of them have two derivations:

"elle essaie des chapeaux de dame"
"j'ai des pas de jouets"
"je veux du jus d'orange"

Rule 10.3 SN** -> EP NP* CC NP* 3 types (4 tokens)

"a Madeleine sans Josiane et Ginette"
"c'est pour les monsieurs et les dames"
"je veux de l'eau et une timbale"

Rule 10.4 SN** -> EP NP* CC EP NP* 5 types (5 tokens)

"cette voiture-là elle veut boire du vin
et de l'eau aussi"
"y a des wagons et du train"

Rule 10.5 SN** -> EP EP NP* 30 types (31 tokens)

In the utterances recognized by this rule, the first EP is a preposition in the classic sense, the second one is du, de, des:

"je joue avec des voitures"
"le métro il a des roues avec des pneus"
"pour travailler à des meubles"
"il va jouer avec du sable"

Rule 10.6 SN** -> EP PT 24 types (29 tokens)

"c'est à moi"
"c'est pour moi"
"j'ai joué avec eux"
"non je veux pas aller avec toi"
"parce que je veux pas aller chez elle"

Rule 10.7 SN** -> EP PR 30 types (39 tokens)

All the utterances recognized by this rule are questions:

"c'est chez qui là ?"
"c'est à qui ce fauteuil ?"
"dans quoi tu vas le mettre ?"
"on peut faire des quoi ?"
"à quoi ça sert une fronde ?"

III.11 Group 11 Determiner Introduction

This group of rules contains the detailed structure of the noun phrases. They determine the way in which determiners, numerical adjectives, adjectives in pre- and post-position can be generated. They are comparable to the rules used in the analysis of noun phrases in Philippe I [9] but they are more detailed and used more specifically. While in our previous study of noun phrases the same rules could recognize proper nouns and common nouns alike, the present grammar recognizes proper and common nouns with distinct rules. In effect, noun-phrases including proper nouns have a more simple structure than noun phrases formed with common nouns. An example that illustrates this point is the rule:

NP* -> NUM ADJP NP' POSTADJP

where NP' can be rewritten as NP (proper noun) as well as NC (common noun). This rule recognizes:

(1) AC AQ NP AQ "deux gentils papa blonds"

(2) AC AQ NC AQ "deux petites cabines bleues"

While (1) is very unlikely, even in adult language, (2) is quite likely and actually occurs in the corpus.

On the basis of this observation, we have written distinct rules for generating complex noun phrases with proper nouns and noun phrases with common nouns.

Rule 11.1 NP* -> NUM ADJP NC POSTADJP 1 type (1 token)

"chercher deux petites cabines bleues"

Rule 11.2 NP*-> DET ADJP NC POSTADJP 12 types (14 tokens)

"après il a mangé le petit chaperon rouge"
"une petite fenêtre ronde"

Rule 11.3 NP* -> ADJP NC POSTADJP 1 type (1 token)

"mais je prépare des petites cabines bleues"

Rule 11.4 NP* -> DET NUM NC POSTADJP 2 types (2 tokens)

"c'est quoi ces deux trucs rouges dessus?"
"le premier wagon rouge ça fait trois"

Rule 11.5 NP* -> NUM NC POSTADJP 2 types (2 tokens)

"trois bougies rouges"
"a deux fils électriques"

Rule 11.6 NP* -> DET NC POSTADJP 51 types (90 tokens)

"un feu rouge"
"une feuille blanche"
"un camion drôle"

Rule 11.7 NP* -> NC POSTADJP 14 types (28 tokens)

"bébé sage"
"pieds nus"
"crayon rouge"

Rule 11.8 NP* -> DET NUM ADJP NC 1 type (1 token)

"je veux voir les deux petits trucs là"

Rule 11.9 NP* -> NUM ADJP NC 7 types (7 tokens)

"deux petits disques"
"deux petites filles elles s'appellent"

Rule 11.10 NP* -> DET ADJP NC 197 types (355 tokens)

"un petit lapin"
"un pauvre garçon"

Rule 11.11 NP* -> ADJP NC 60 types (133 tokens)

"grosse moto"
"nouvelle maîtresse"

Rule 11.12 NP* -> DET NUM NC 7 types (8 tokens)

"les deux ballons"
"le second ballon"

This rule also recognizes un deux trous which should not be parsed. This is one more case where the punctuation would have improved the grammatical analysis.

Rule 11.13 NP* -> NUM NC 32 types (54 tokens)

"je prends deux cuillères"
"Ginette m'a donné deux gâteaux"
"Trois heures"

In most cases, the terminal word for ac is "deux".

Rule 11.14 NP* -> DET NC 1264 types (3338 tokens)

"un chapeau"
"la fanfare"
"mon croissant"
"de quelle couleur?"
"regarde ce feu"

Rule 11.15 NP* -> NP" 1064 types (2640 tokens)

As we shall see, NP" is rewritten as NC, NP, DN PO, PI, DET PI. Twenty percent of the utterances recognized by this rule are one- or two-word utterances.

246 NC caca, pipi, raquette, bateau.

154 EP NC par terre, du feu, des animaux, à pied, au zoo,

de côté

137 NP Papa, Maman, Madeleine, Philippe

Rule 11.16 NP* -> DET AI NC 38 types (72 tokens)

"une autre tige"
"un autre train"

In most cases, the terminal word for AI is autre. In a few cases it is même:

"le même camion"

Rule 11.17 NP* -> AI DET NC 27 types (38 tokens)

"tous les crayons"
"regarde toutes ces voitures"
"j'ai pas mangé tout mon croissant"

Rule 11.18 NP* -> AI DET NC AD 2 types (4 tokens)

Rule 11.17 and 11.18 perform the same analysis, the only difference being that rule 11.18 recognizes compound forms of the demonstrative adjective ce....-là:

"c'est quoi toutes ces roues-là?"
"tous ces puzzles-là"

Rule 11.19 NP* -> AI NC 31 types (37 tokens)

"ça dit pas quelque chose ?"
"quelque chose qui est un peu noir"
"ils sont de chaque couleur"
"a des autres roues"

Rule 11.20 NP* -> DET NUM 9 types (15 tokens)
"les deux"
"une première"
"j'en ai un troisième"

Rule 11.21 NP* -> DET NP 28 types (34 tokens)

"de ta Maman"
"quelle Liliane ?"
"c'est mon Papa"

Rule 11.22 NP* -> DET ADJP NP 4 types (4 tokens)

"mon gentil Papa"
"qu'est-ce qu'il va faire le petit Indien?"

Rule 11.23 NP* -> ADJP NP

"vilaine Madeleine"
"petit Pascal"

Out of the 13 types (25 tokens) recognized by this rule, 7 types (18 tokens) have two distinct trees, rule 11.23 being correct only when the adjective modifies the proper noun. In the other cases, NP is a vocative:

"c'est dur Maman"
"c'est beau Clara"

Here again, punctuation would facilitate the analysis.

Rule 11.24 NP* -> DET NC AD 50 types (72 tokens)

This rule is analogous to rule 11.14, the only difference being that it recognizes compound forms of the demonstrative adjective:

"c'est quoi cette bête-là?"

Rule 11.25 NP* -> DET ADJP NC AD 5 types (6 tokens)

This rule is analogous to rule 11.10, and it recognizes the compound forms of the demonstrative adjective:

"c'est quoi ce gros truc-là?"

III.12 Group 12 Noun-Phrase Utterances

As we have already said, SN*1 is a non-terminal symbol which identifies noun phrases that occur as a complete utterance, without any

verb. The rules of this group replicate the rules of groups 9 and 10, but all of them have not been reproduced in group 12.

Since most of these rules have already been commented upon when presenting Groups 9 and 10, we shall give less details.

Rule 12.1 SN*1 -> NP* 99 types (1385 tokens)

Most of the utterances recognized by this rule have two words:

"une moto"
"la musique"

Three types have only one word:

246 NC canard, cheval, hélicoptère

137 NP Papa, Maman, Babar, Saint-Cloud

6 PI rien

The longest utterances have five words; there are only 5 of them:

"et pas deux petits trous"
"et un petit nez noir"
"un grand grand grand plateau"
"et une petite bouche aussi"
"non un gros camion dedans"

Rule 12.2 SN*1 -> NP* CC NP* 9 types (17 tokens)

"la pelle et le seau"
"la tortue et le chat"

Rule 12.3 SN*1 -> NP* EP NP* 45 types (161 tokens)

"un camion de pompiers"
"un filet à papillons"
"un grain de riz"
"le bateau de papa"

Rule 12.4 SN*1 -> NP* EP EP NP* 5 types (8 tokens)

"une canne avec des dessins"
"un magasin pour des musiques"

Similarly to Rule 9.5, Rule 12.4 recognizes utterances in which there is a repetition of the EP:

"et les mains de au le monsieur"
"et les bouteilles de d'eau"

Rule 12.5 SN*1-> NP* CC EP NP* 1 type (2 tokens)

"une timbale et de l'eau"
"un bateau et de l'eau"

Rule 12.6 SN*1 -> NP* EP PR 3 types (6 tokens)

"le rocher de qui ?"
"le fauteuil de qui ?"
"une peau de quoi ?"

Rule 12.7 SN*1 -> NP* EP PT 6 types (8 tokens)

"une petite cuillère pour moi"
"voilà un tableau pour toi"

Rule 12.8 SN*1 -> PF 5 types (19 tokens)

"et moi"
"pas moi"
"non toi"

Rule 12.9 SN*1 -> PD 7 types (145 tokens)

"ça"
"celui-là"

Rule 12.10 SN*1 -> PD EP NP* 4 types (8 tokens)

This rule was not included in Group 9. It has been introduced to parse utterances of the following form:

"celle de Gérard"
"celle-là pour Philippe"
"celui de Papa"
"celle-là dans le tiroir"

Rule 12.11 SN*1 -> EP NP* 49 types (521 tokens)

"des animaux"
"de l'eau"
"sur les rails"

"à la télévision"
"sans la cuillère"

Rule 12.12 SN*1 -> EP NP* EP NP* 18 types (44 tokens)

"des roues pour l'avion"
"au bord de la table"
"des livres dans le tiroir"
"de l'eau sur la table"

Rule 12.13 SN*1 -> EP NP* CC NP* 4 types (6 tokens)

"sur les fusils et les pistolets"
"avec une culotte et un chapeau"

Rule 12.14 SN*1 -> EP NP* CC EP NP* 3 types (4 tokens)

"des petites voitures et des maisons"
"pour Philippe et pour Maman"

Rule 12.15 SN*1 -> EP NP* EP EP NP* 3 types (4 tokens)

"des boutiques avec des jouets"
"du metal avec des moules"

Rule 12.16 SN*1 -> EP EP NP* 8 types (19 tokens)

"avec de la farine"
"à des gens"
"sur des croissants"

Rule 12.17 SN*1 -> EP PT 2 types (18 tokens)

"pour moi"
"à moi"
"de toi"

Rule 12.18 SN*1 -> EP PR 1 type (15 tokens)

"avec quoi?"
"à qui ?"
"dans quoi?"
"de qui ?"

Rule 12.19 SN*1 -> EP PD 1 type (8 tokens)

This rule does not appear in Group 10. It parses utterances of the following form:

"dans celle-là"
"avec ça"
"pour celui-là"
"sur celle-là"

III.13 Group 13 Some Pronouns and Common and Proper Nouns

This group includes the rewrite rules for NP", a non-terminal symbol which has been introduced in Group 11.

Originally this group contained personal and demonstrative pronouns. In an attempt to improve the probabilistic fit, personal and demonstrative pronouns, which occur more frequently than possessive and indefinite pronouns, have been entered at higher levels, and are generated through specific rules. In our first attempts they were not generated by distinct rules.

Rule 13.1 NP" -> DN PO 11 types (17 tokens)

DN PO is the terminal symbol for possessive pronouns:

"le mien"
"celui-là c'est le mien"
"je prends le tien"

Rule 13.2 NP" -> PI 231 types (279 tokens)

PI is the terminal symbol for indefinite pronouns. In 64 percent of the types, PI is the subject of the verb:

"on le pèse"
"on le casserait"
"comment on les fabrique les bols?"

When PI is not the subject, the terminal words for it are mainly: plusieurs, rien, tout, tous.

"il les a enlevés tous"
"elle m'a dit rien"

Rule 13.3 NP" -> DET PI 35 types (71 tokens)

This rule is also a rewrite rule for indefinite pronouns: those which have a compound form.

"c'est le même"
"fais un autre"

Rule 13.4 NP" -> NP 329 types (839 tokens)

As we already said, NP is the terminal symbol for proper nouns.

Rule 13.5 NP" -> NC 551 types (1456 tokens)

NC is the terminal symbol for common nouns.

III.14 Group 14 Adjective Phrases

The notation ADJP is introduced for adjective in pre-position.

(3) AQ is the terminal symbol for qualitative adjectives.

14.1 ADJP -> AQ 299 types (545 tokens)

"c'est pas une bonne idée"
"c'est un gros mouton ça"

14.2 ADJP -> ADJP AQ 8 types (9 tokens)

"dernier petit trou"
"où il est le grand méchant loup"

In the utterances recognized by our grammar there is only one case where there are more than two adjectives in succession:

"un grand grand grand plateau"

(3) Pre-position refers to adjectives that precede the noun modified, such as bon; post-position refers to adjectives that follow the noun, such as facile. Of course some adjectives may be either, with variations of meaning.

It should be mentioned that this rule recognizes the following utterance, which is not correctly analyzed:

"cochon sales mains"

III.15 Group 15 Post-position Adjectives

The rule which generated 2 or more adjectives in post-position, in the grammar for the noun phrases, has been eliminated, since it was not needed for Philippe's speech.

15.1 POSTADJP -> AQ 83 types (138 tokens)

"ton manteau blanc"
"c'est une moto anglais"
"c'est un jeu amusant"

III.16 Group 16 Determiners

This group permits rewriting the non-terminal symbol DET as any of the five terminal symbols standing for a grammatical category.

16.1 DET -> AD 146 types (234 tokens)

AD is the terminal symbol for demonstrative adjectives

16.2 DET -> AT 16 types (32 tokens)

AT is the terminal symbol for interrogative adjectives

16.3 DET -> AO 204 types (335 tokens)

AO is the terminal symbol for possessive adjectives

16.4 DET -> DN 992 types (2418 tokens)

DN is the terminal symbol for definite articles

16.5 DET -> IN 425 types (1205 tokens)

IN is the terminal symbol for indefinite articles in the singular form only. Note that indefinite articles in the plural form are coded EP; consequently, any comparison between the number of types recognized by rules 16.4 and 16.5 would be misleading.

III.17 Group 17 Numerical Expressions

NUM is the non-terminal symbol for numerical expressions. AC is the terminal symbol for cardinal adjectives. AN is the terminal symbol for ordinal adjectives.

Rule 17.1 NUM -> AC 59 types (114 tokens)

Rule 17.2 NUM -> AN 5 types (5 tokens)

Before presenting the rewrite rules of the verb phrase, we shall note again that the present analysis does not take into account the mode, the tense or the person of the verb, since subscripts which convey this sort of information are disregarded by the grammar. As a result, our grammar recognizes sentences in which the verb has a correct form as well as sentences in which the mode, the tense, or the person is not appropriate. For example,

"elle enlève ses lunettes"
"il traire sa vache"

are both recognized.

Our definition of the grammaticality of a sentence produced by Philippe is given by reference to the structure of the sentence, and

does not depend on the choice of the individual words within the sentence or their appropriateness to convey the intended meaning. Our grammar does not recognize an utterance if its elements are not arranged in a certain order, but allows, in certain cases, for elements of a classical French sentence to be missing.

For instance, while la fumée on s'amuse la fumée is not recognized, noun phrases in which the determiner is missing, or sentences in which a personal pronoun is missing are both recognized.

"gros collier" (DET is missing)
"fume la maîtresse" (PF is missing)
"veux du médicament" (PF is missing)
"Papa parti" (Auxiliary is missing)

Thus the "constituants obligatoires" of the noun phrase (SN) and of the verb phrase (SV) are minimized beyond what is standard French. The basic rewrite rules of SN and SV can be summarized as follows:

```
SN ->  DET  N

SV ->  AUX  GV
GV ->  V SNP
GV ->  COP SN
         SA
         SNP
```

where DET is the determiner, N is the noun (common or proper), AUX is the auxiliary, V is the verb, SNP is the noun-phrase object of the verb or a prepositional noun phrase, COP is the copula, SN is the nominative predicate, and SA the adjectival predicate.

On the basis of this general scheme, we have introduced several groups of rules which allow the parsing of different forms of verb phrases.

Originally, following Dubois' suggestion [1] (p.93), AUX was rewritten as an auxiliary (être and avoir) and as a moda'. With the intended semantic interpretation in mind, AUX has been divided into two categories: AUX 1 which stands for the auxiliary verbs être and avoir (VA), and AUX 2 which stands for modals (VD).

Contrary to Dubois, AUX is not a "constituant obligatoire" of the verb phrase, since mode, tense and person which, for Dubois, constitute an element of the verb phrase, are not taken into account. If we had written AUX as a "constituant obligatoire" of the verb phrase, we could not have parsed utterances in which the auxiliary is missing.

We have distinguished two forms of the verbal group: GV and GV'. While in the rewrite rules of GV the first element on the right-side of the arrow is a verb or the non-terminal symbol PY which stands for "y", followed by a verb, or the adverb of negation "ne" followed by a verb, in the rewrite rules of GV', the first element on the right-side is always a personal pronoun (direct or indirect object) or a preposition.

III.18 Group 18 Verb Phrase Structures

We shall now present the rules of the verb phrase (SV).

Rule 18.1 SV -> GV 1667 types (3997 tokens)

Three types have only one verb:

109 VT
48 VN
18 VI

41 types (281 tokens) have only two words:

26 PF VN	4 DG VN
21 VT PD	4 KO VN
20 PF VT	4 KO VT
18 VT NP	4 VN DG
16 PI VN	3 DV VI
16 VN NP	3 KO VI
14 PD VI	3 PF VC
14 VT PT	3 VI DV
10 PF VI	3 VN PD
9 VT NC	3 VT PI
8 VN NC	3 NP VI
8 VT DG	2 PD VT
7 PD VN	2 PY VT
7 PI VT	2 VI DG
7 VT DV	2 VI NP
6 DV VN	2 VT KO
6 VN DV	1 CC VN
6 VN PT	1 DG VI
5 DG VT	1 NP VN
5 DV VT	1 PF VM
	1 VT PR

Rule 18.2 SV -> GV' 296 types (584 tokens)

As already mentioned, group GV' gives the rewrite rules of verb phrases which begin with a personal pronoun (direct or indirect object) or with a preposition

4 types (26 tokens) have only two words:

9 EP VT
7 EP VN
5 PS VN
5 PS VT

"je le défais"
"s'il te plaît"
"pour lancer la balle"
"en a deux"
"pour accrocher au vélo"

Rule 18.3 SV -> AUX1 GV 263 types (476 tokens)

This is the basic rule for rewriting compound forms of the verbs:

"il a perdu sa queue"
"il est sorti de l'eau"

Rule 18.4 SV -> AUX1 DV GV 6 types (10 tokens)

This rule allows the insertion of an adverb between the auxiliary and the verb:

"j'ai presque fini"
"non j'ai trop mangé"
"il a déjà fondu le gros morceau de sucre"

It gives an incorrect analysis of the following utterance:

"devine qu'est-ce y a dedans devine"

which is coded VT PR,DV,CS LC PY VT,VA DV VT and recognized as VT PR LC PY VA DV VT.

Rule 18.5 SV -> AUX1 DG GV 22 types (33 tokens)

DG is the terminal symbol for adverbs of negation. This rule recognizes compound forms of the verb in the negative form when the negation is incomplete, i.e., the first element of the negation "ne" is omitted, as usual in spoken French:

"j'ai pas vu"
"j'étais pas couché"

Rule 18.6 SV -> AUX2 GV' 90 types (167 tokens)

AUX2 is the non-terminal symbol for the modals, those verbs which express the idea of permission, willingness, possibility, impossibility:

"je vais te renvoyer de l'eau"
"il faut le tourner"
"je vais la mettre là la remorque"
"je veux l'écouter encore un peu"

Rule 18.7 SV -> AUX2 DG GV' 23 types (33 tokens)

This rule introduces the negation which applies to the modal:

"je veux pas la voir Madeleine"
"il faut pas l'allumer"
"on peut pas le voir"
"je veux pas te le dire"

Rule 18.8 SV -> AUX2 GV 253 types (383 tokens)

"faut racheter des gâteaux"
"va monter sur le trottoir"

Rule 18.9 SV -> AUX2 DV GV 6 types (6 tokens)

An adverb which modifies the modal can be introduced by this rule:

"on peut aussi nager dans l'eau"
"je vais bientôt aller chez Mamie"

Rule 18.10 SV -> AUX2 DG GV 33 types (43 tokens)

This rule recognizes sentences in which the negation applies to the modal:

"veux pas manger"
"il veut pas passer"
"tu peux pas rentrer"
"je veux pas voir Ginette"

Rule 18.11 SV -> PS AUX1 GV 91 types (118 tokens)

This rule parses utterances in which a personal pronoun that is the object of the compound verb appears before the verb:

"tu les as emportés les croissants"
"il les a enlevés tous"
"je l'ai cachée ma tête"
"je l'ai attrapé"

Rule 18.12 SV -> PS AUX1 DG GV 10 types (16 tokens)

This rule recognizes the same sorts of utterances as Rule 18.11, but the sentence is in the negative form:

"non je l'ai pas vue"
"on l'a pas attrapé"
"je les ai pas comptés"

Rule 18.13 SV -> PS AUX2 GV 6 types (6 tokens)

This rule recognizes utterances in which the personal pronoun is the direct object of the verb which follows the modal:

"tu prends du chocolat avec le pain et tu le laisses couler"
"je veux que tu le fasses marcher un coup"

This rule recognizes a sentence which should not have been parsed:

"tu me viens voir comment je fais un petit bateau"

Rule 18.14 SV -> PS PS AUX1 GV 4 types (5 tokens)

This rule recognizes sentences in which the first personal pronoun is the indirect object and the second one the direct object of the verb:

"c'est Ginette qui te l'a donné"
"on me l'a nettoyé"

Rule 18.15 SV -> PY AUX1 GV 10 types (13 tokens)

This rule recognizes sentences in which "y" precedes the verb in the compound form:

"là y a écrit soleil"
"là y a marqué tissus"
"on y est allés un jour"
"qu'est-ce qu'y a écrit?"

III.19 Group 19 Auxiliaries

Rule 19.1 AUX1 -> VA 395 types (660 tokens)

VA is the terminal symbol for être and avoir when they are used as auxiliaries.

Rule 19.2 AUX1 -> VA VD 11 types (11 tokens)

This rule recognizes sentences in which the modal has a compound form:

"mais si tu as fait chauffer"
"c'est qui a fait tomber ces cigarettes ?"
"a pu sortir"

III.20 Group 20 Modals

Rule 20.1 AUX2 -> VD 392 types (616 tokens)

VD is the terminal symbol for the modals:

"ça va parler"
"je veux me frotter les dents"
"je veux voir les photos"

Rule 20.2 AUX2 -> VD VD 10 types (12 tokens)

This rule recognizes sentences in which there are two modals in succession, which express the notion of future:

"je vais aller galoper"
"on va aller la voir"
"je vais faire tomber ton briquet"
"je veux aller jouer"

Rule 20.3 AUX2 -> VD PS VD 7 types (8 tokens)

In all the utterances parsed, the terminal word for the second VD is faire:

"faut me faire manger"
"pourquoi ils vont les faire arrêter ?"
"il va se faire écraser"
"on va les faire brûler avec du feu"

III.21 Group 21 Verbal Groups

As already mentioned, GV is the non-terminal symbol for verbal group. In addition, COP (copula) and V (verb) are essential non-terminal symbols for the verbal group.

COP is rewritten as VC. We followed Dubois' suggestion [1] (p. 79) to code as VC only the three following verbs: être, devenir, rester; the other verbs of state (paraître, sembler) that traditional grammars classify within the same category as être are coded VI (intransitive verbs). Etre is also coded VA, in order to take into account its role as an auxiliary. Devenir and rester are multiply coded VI as well as VC.

COP can be followed by an adjectival phrase (SA), or a noun phrase (SN*) or a prepositional noun phrase (SN**), or an adverb (DV), or an adverbial locution (LC).

Rule 21.1 GV -> COP 23 types (76 tokens)

We have introduced this rule in order to recognize sentences in the interrogative form which begin with an adverb of interrogation:

"où il est?" DV PF VC

Such sentences are parsed by using the following rules:

S
1.3 -> S3
4.3 -> DV S'

5.2 -> DV SNS SV
7.2 -> DV PF SV
18.1 -> DV PF GV
21.1 -> DV PF COP
24.1 -> DV PF VC

The other types of questions recognized by rule 21.1 are:

"comment il est?"
"mais où il est?"

This rule also recognizes sentences in which the prepositional noun phrase or the nominative predicate are before the verb:

"en Angleterre elle est"
"losange c'est" (for "c'est un losange")

A wrong analysis is obtained for the following utterance:

"fou il est" (for "il est fou")

where "fou" is analyzed as NC instead of AQ.

Rule 21.2 GV -> COP LC 5 types (8 tokens)

This rule recognizes sentences in which an adverbial locution follows the verb être:

"il est là-bas"
"ah c'était là-dedans"
"qu'est-ce c'est là-dedans?"

Rule 21.3 GV -> COP SA 45 types (193 tokens)

Adjective-predicates are parsed by this rule:

"c'est chaud"
"c'est dur"
"c'est très drôle"

Rule 21.4 GV -> COP SN 250 types (666 tokens)

Two types of noun phrases are parsed by this rule:

- Nominative predicate:

"c'est papa"

"c'est une dame"
"c'est des roues"

- Prepositional noun phrases:

"ils sont dans la maison"
"c'est à moi"
"là c'est chez Mamie"

Rule 21.5 GV -> COP SA SN 34 types (66 tokens)

In sentences parsed by this rule, the noun phrase SN emphasizes the pronominal subject of COP:

"il est beau cet hélicoptère"
"il est gros ce bol"

This rule parses also sentences in which the pronoun subject has been omitted:

"est gourmand ce Papa"
"est difficile ça"

Rule 21.6 GV -> COP SA DV 3 types (3 tokens)

This rule recognizes the three following utterances:

"c'est joli aussi"
"c'est trop chaud encore"
"elle est toute dure maintenant"

Rule 21.7 GV -> COP SN DV 23 types (29 tokens)

This rule is analogous to rule 21.4, the only difference being the adverb at the end of the sentence:

"c'est chez qui là?"
"qui c'est ce bébé là?"
"que c'est ça dedans?"
"elle est dans ma chambre là"

Rule 21.8 GV -> COP DV SN 27 types (43 tokens)

This rule recognizes sentences in which an adverb separates the copula and the nominative predicate:

"il est là le moteur"
"il est dedans le conduiteur"
"c'est là le ciel"
"il était là sur le papier"

Rule 21.9 GV -> COP DG SA 16 types (47 tokens)

This rule generates negative sentences with an adjective predicate:

"c'est pas vrai"
"c'est pas possible"
"c'est pas très bien"
"parce qu'elle est pas belle"

Rule 21.10 GV -> COP DG SN 34 types (48 tokens)

This rule generates negative sentences with a nominative predicate or a prepositional noun phrase:

"non c'est pas la même"
"c'est pas une bonne idée"
"c'est pas à Mamie"
"c'est pas chez Mamie"

Rule 21.11 GV -> COP dv 22 types (52 tokens)

This rule recognizes sentences in which the verb <u>être</u> is followed by an adverb:

"je suis là"
"le ciel c'est là"

In a few utterances, the verb is <u>rester</u>:

"Maman reste là"
"non je veux rester ici"

<u>Rester</u> is coded VC,VI. The grammar recognized both VC DV and VI DV; probabilistic disambiguation eliminates VI DV.

Rule 21.12 GV -> COP DG DV 5 types (8 tokens)

This rule is analogous to rule 21.11, but it introduces the negative particle DG, as in:

"il est pas là"
"Misette est pas là"

Rule 21.13 GV -> COP GV' 22 types (28 tokens)

This rule recognizes sentences in which the verb etre is followed by pour:

"c'est pour faire quoi ces boutons ?"
"ça c'est pour monter"
"c'est pour l'allumer"
"c'est pour l'arrêter le gros bouton"

This rule analyzes incorrectly the two following sentences. Their meaning is perfectly clear, but they are not well constructed:

"ils sont en a encore là" (Philippe stuttered)
"qui c'est en veut du croissant?"

Rule 21.14 GV -> COP DG SA SN 17 types (23 tokens)

This rule is analogous to rule 21.5, but it analyzes sentences in the negative form:

"elle est pas gentille Madeleine"
"non c'est pas bon les gouttes dans le nez"
"c'est pas joli son vêtement"

Rule 21.15 GV -> COP SN SNS 94 types (132 tokens)

This rule recognizes five sorts of sentences:

a) sentences in which the nominative predicate is followed by ca:

"c'est un manège ça"
"c'est le pont ça"
"c'est le nom de quoi ça?"

b) sentences in which the nominative predicate is a pronoun, the noun-phrase which follows it being in apposition to the neutral subject c':

"c'est ça ton sac"
"c'est celui-là le bras gauche"

c) sentences in which the first noun phrase is prepositional, and the second noun phrase is in apposition to the subject:

"il est au cabinet Marc"
"elle est sur le camion la brosse"

d) sentences in which the second noun phrase modifies the first noun phrase and in this case the grammatical analysis in incorrect:

"c'est un camion bétonneuse"

e) Finally, this rule recognizes also the following utterances and the grammatical analysis is incorrect.

"c'est papa maman"
"c'est Gérard Martine"
"c'est Babar Babar"

Rule 21.16 GV -> COP DV SN DV 3 types (5 tokens)

The following forms of sentences are recognized:

"ça c'est encore chez Mamie là"
"ça c'est encore une fenêtre là "

Rule 21.17 GV -> COP DG SN GV' 2 types (3 tokens)

The group GV' is the complement of the noun phrase:

"c'est pas des papiers pour jouer"

An idiomatic form is also recognized by this rule:

c'est pas la peine de (it is not worthwhile):

"c'est pas la peine de le défaire"

Rule 21.18 GV -> COP DV' AQ 3 types (4 tokens)

This rule parses sentences in which the adjective phrase is modified by the expression "un peu" which is classified as an adverb:

"quelque chose qui est un peu noir"
"et j'ai été un petit peu malade"
"voilà les gens ils sont un peu idiots"

Rule 21.19 GV -> COP DG DV SN 3 types (9 tokens)

The following forms of sentences are recognized:

"elle est pas là la machine"
"non elle est pas toujours malade"

Rule 21.20 GV -> COP SN GV' 21 types (21 tokens)

In this rule GV' is the complement of the noun phrase:

"c'est un petit savon pour laver les dents"
"c'est un plateau pour mettre les dés"

This rule also parses the expression en train de faire (I am doing, he is doing):

"je suis en train de manger une pomme"
"qu'est que tu es en train de faire Marc?"

Rule 21.21 GV -> COP DM V 2 types (4 tokens)

Here the adverb modifies the verb V:

"il est tout défait"

Rule 21.22 GV -> COP DM V SN 2 types (2 tokens)

"maintenant il est tout froissé mon moteur"
"il est tout défait le truc"

Rule 21.23 GV -> PY COP 4 types (16 tokens)

This rule recognizes sentences in which y precedes être:

"ca y est"
"la valise elle y est"

"les cadeaux y sont"

Rule 21.24 GV -> V 283 types (751 tokens)

This rule recognizes verb phrases formed with transitive verbs (VT), and verbs which can be transitive or intransitive (VN) when there is no direct object or prepositional noun phrase.

There are 109 transitive verbs (VT) that occur alone:

"manger"
"regarder"

and 48 transitive/intransitive verbs (VN) which occur alone:

"tenir"
"casser"

This rule recognizes also sentences in which the verb has a compound form:

"il est cassé"

as well as sentences which include a modal:

"je vais voir"

Finally, this rule recognizes sentences in which the direct object of a compound verb is a personal pronoun:

"il les a enlevés"

Rule 21.25 GV -> V SNP 671 types (1466 tokens)

As we already mentioned, SNP is a non-terminal symbol which allows the parsing of noun phrases which are direct objects of verbs, or prepositional noun phrases:

"mange le yaourt"
"fermer les yeux"
"il dirige les voitures"
"mettre dans le tracteur"

In a few cases, however, Rule 21.25 recognizes sentences in which the noun phrase which follows the verb is not the object, but the subject of the verb:

"fume la maîtresse"

Finally, the sentence may be ambiguous:

"tourne la clef"

which, for Philippe, may mean "turn the key" or "the key turns". The same utterance, produced by an adult would not be ambiguous, and would be interpreted as an imperative.

Rule 21.26 GV -> V SNP DV 99 types (136 tokens)

This rule recognizes sentences which have the same structure as those recognized by rule 21.25, but which end with an adverb:

"avait du feu dehors"
"a des gens aussi"
"j'enlève le carrelage aussi"
"je mets les allumettes dedans"

Rule 21.27 GV -> V DV 48 types (91 tokens)

"qu'est-ce y a dedans?"
"qu'est-ce y a encore?"
"l'eau sort là"

Rule 21.28 GV -> V DV SNP 47 types (63 tokens)

"on fait comme ça"
"il a seulement un camion"
"je veux encore du chocolat"

Rule 21.29 GV -> V DG SNP 62 types (101 tokens)

This rule recognizes sentences in the negative form. It should be noticed that the negation is incomplete, as is usual in spoken French, the word "ne" being omitted:

"il a pas de queue"

"non je l'ai pas vue"
"non c'est pas la même"

This rule recognizes also sentences in the negative interrogative:

"ça dit pas quelque chose?"

Rule 21.30 GV -> V DG SNP SN**

We do not present the types and tokens parsed by this rule, since all of them have two derivations. This rule allows a correct analysis of utterances in which the prepositional noun phrase is related to the verb, while rule 21.29 parses correctly noun phrases which are the complement of one other noun phrase.

While rule 21.29 analyzes correctly the following forms:

"on voit pas le nez du monsieur"
"parce qu'il a pas la forme des autres poissons"

rule 21.30 recognizes correctly such sentences as:

"non j'ai pas de zoutils à l'école"
"non a pas des animaux sur le manège"

Rule 21.31 GV -> NE V DG SNP 1 type (1 token)

This rule recognizes sentences in the negative form, when the whole negative formula ne...pas is used:

"ils n'ont pas de siège"

Rule 21.32 GV -> V LC 4 types (9 tokens)

This rule recognizes sentences in which a verb is followed by a locution, whatever its nature:

"je fais semblant"
"qu'est-ce y a là-bas?"

Rule 21.33 GV -> V DG DV 11 types (36 tokens)

"je sais pas pourquoi"
"je sais pas trop"

"non je mange pas vite"

Rule 21.34 GV -> V SNP SNS

We do not present the number of types and tokens analyzed by this rule, since most of them have two derivations. The aim of this rule is to analyze sentences in which the subject is repeated at the end of the sentence:

"fait du bruit la voiture"

Rule 21.35 GV -> V SNP SN**

We also do not present the number of types recognized by this rule, since each type has two derivations. This rule parses utterances containing a direct object followed by an indirect objet, such as:

"mettre du feu sur la cigarette"
"mets des valises dans le car"

and recognizes also sentences of the following form, where the preposition "de" indicates the possessive case:

"mettre les lunettes de Madeleine"

These types are correctly parsed by rule 21.25, as we explained when presenting rule 9.3.

Rule 21.36 GV -> V DG 35 types (178 tokens)

This rule analyzes verb phrases in the negative form:

"ça roule pas"
"non il fume pas"

Rule 21.37 GV -> V PT 9 types (29 tokens)

This rule parses two forms of utterances:

a) imperative form:

"montre-moi"
"cache-le"
"lève-toi"

b) it gives an incorrect analysis of sentences in which the personal pronoun that is at the end of the sentences repeats the subject of the verb:

"je veux pas boire moi"
"veux tourner moi"

Rule 21.38 GV -> V PT V 1 type (3 tokens)

This rule has been introduced to parse the following form:

"laisse-moi faire"

Rule 21.39 GV -> V PT SNP 12 types (17 tokens)

This rule analyzes utterances in which the imperative is followed by a direct object, or a prepositional noun phrase:

"donne-moi la voiture"
"remets-la sur le camion"

Rule 21.40 GV -> V PT PT 2 types (3 tokens)

This rule recognizes sentences in which the second personal pronoun is the indirect object of the verb:

"défais-le moi"
"montre-la moi"

Rule 21.41 GV -> V PT DV 1 type (4 tokens)

This rule is introduced to parse the imperative form of the verb s'en aller:

"va-t'en"

Rule 21.42 GV -> V PT DV SNP 1 type (19 tokens)

As rule 21.41, this rule is "ad hoc" and parses sentences of the following form:

"va-t'en la fumée"

Rule 21.43 GV -> V DV' 3 types (3 tokens)

This rule analyzes sentences in which the expression un peu or un petit peu plays the role of an adverb:

"non je veux pas manger un peu"
"j'ai pleuré un petit peu"
"j'ai renversé un petit peu"

Rule 21.44 GV -> V DV DV' 1 type (2 tokens)

"je veux écouter encore un peu"

Rule 21.45 GV -> V DV' SN 8 types (11 tokens)

The rule recognizes three sorts of sentences:

a) sentences in which the expression un peu or un petit peu is followed by a noun phrase:

"moi veux un petit peu de lumière"
"elle fait un petit peu de bruit"

b) sentences in which the final noun phrase repeats the subject:

"elles roulent un petit peu les voitures"

c) sentences in which the final noun phrase is prepositional:

"je pleure un peu à l'école"

Rule 21.46 GV -> V DV DV' SN 1 type (1 token)

"je veux boire encore un peu d'eau"

Rule 21.47 GV -> V DM DV 2 types (2 tokens)

In the sentences parsed by this rule, the first adverb modifies the second one:

"et puis ça ça roule très loin"
"il va rouler très vite"

Rule 21.48 GV -> V DM AQ 15 types (20 tokens)

"je mange tout seul"
"ah elle roule tout seul"
"mon moteur il a sauté tout seul"

Rule 21.49 GV -> V NUM 4 types (4 tokens)

"le premier wagon rouge ça fait trois"
"j'en ai pris deux"

Rule 21.50 GV -> V DET AQ 6 types (8 tokens)

This rule parses sentences in which the substantive direct object of the verb is omitted:

"pas fait un gros"

Rule 21.51 GV -> V EP AQ 5 types (7 tokens)

"le crayon qui dessine du marron"
"j'ai du vert"
"regarde qu'est-ce je vais faire de beau"

Rule 21.52 GV -> V SNP PR S'

This rule has been introduced to parse sentences in which a verb in the imperative form is followed by a sentence, as in:

"regarde ce que je vais cacher"
"regarde ce que j'ai fait Madeleine"

This same rule parses utterances of the following form:

"non a le gentil loup qui va le défendre"
"et puis a des bateaux qui a des roues"

in which the first verb a is said for the whole expression il y a. Other forms of sentences analyzed by this rule have two derivations and are correctly analyzed by rule 5,5 (S' -> SNS SV PR SV).

Rule 21.53 GV -> V' 67 types (157 tokens)

As we already mentioned, V' is a non-terminal symbol for

intransitive and impersonal verbs. We found it useful to have the same non-terminal for these verbs that cannot have a direct object.

"est tombé"

Rule 21.54 GV -> V' SNS 42 types (67 tokens)

This rule parses the sentences in which the intransitive verb is followed by a noun phrase which is its subject:

"partie la roue"
"est tombé moi"

or in which the subject is repeated:

"où elle va Maman?"

Rule 21.55 GV -> V' DV 19 types (22 tokens)

"il pleut dehors"
"Michel dormir là"
"il va là"

Rule 21.56 GV -> V' SN** 64 types (105 tokens)

This rule recognizes sentences in which the intransitive verb is followed by a prepositional noun phrase:

"on va sur la chaise"
"il va dans l'eau"
"il est tombé dans le chocolat"

Rule 21.57 GV -> V' DG 17 types (22 tokens)

This rule recognizes intransitive and impersonal verbs in the negative form:

"c'est quoi qui marche pas?"
"j'arrive pas"
"[illegible] pleut pas"

Rule 21.58 GV -> V' DG SNS 8 types (11 tokens)

This rule parses sentences in the negative form where the noun phrase is a repetition of the subject:

"elle vient pas Ginette"
"j'arrive pas moi"

Rule 21.59 GV -> V' DV' 3 types (3 tokens)

"faut galoper un petit peu"
"radotes un peu"
"je veux que tu restes un peu"

Rule 21.60 GV -> V' DV' DV 1 type (1 token)

"non non je veux que tu restes un peu encore"

Rule 21.61 GV -> V' DV DV' 2 types (4 tokens)

"moi je tousse encore un peu"

Rule 21.62 GV -> V' DM AQ 4 types (5 tokens)

"et le livre qui tombe tout seul"
"non elle est partie toute seule"

Rule 21.63 GV -> PY V 5 types (7 tokens)

This rule, as well as the following ones, recognizes sentences that contain the word y, which can be an adverb or a pronoun, or enter in the expression il y a.

"qu'est-ce qu'y a?"
"faut pas y toucher"

Rule 21.64 GV -> PY V SNP 31 types (55 tokens)

In the sentences parsed by this rule, the word y is part of the expression il y a:

"y a du vent"
"y a une hélice"

Rule 21.65 GV -> PY V DV 2 types (5 tokens)

"y a dedans"

Rule 21.66 GV -> PY V LC 2 types (5 tokens)

"y a là-haut"

Rule 21.67 GV -> PY V SNP PR SV 11 types (11 tokens)

"y a le chat qui attend"
"y a des trucs qui vont dans tous les sens"
"y a un monsieur qui chante là"

Rule 21.68 GV -> PY V DG 1 type (1 token)

"non y a pas"

Rule 21.69 GV -> PY V DG SNP 6 types (7 tokens)

"parce qu'y aura pas la lumière"
"y a pas le feu"

Rule 21.70 GV -> PY V DV SNP 3 types (9 tokens)

"y a aussi des livres"
"y a dedans dans la vitre"

Rule 21.71 GV -> PY V SNP DV 10 types (18 tokens)

"y avait des belles autos aussi"
"y a sa moto là"

Rule 21.72 GV -> PY PS V DV 3 types (8 tokens)

"y en a encore"
"non y en a beaucoup"

Rule 21.73 GV -> PY PD V DV SNP 2 types (3 tokens)

"y en a beaucoup de ballons"

Rule 21.74 GV -> PY PS V DG 1 type (1 token)

"y en a pas"

Rule 21.75 GV -> PY PS V DG SNP 1 type (1 token)

"y en a pas d'eau"

Rule 21.76 GV -> V PR LC PF SV 10 types (18 tokens)

"regarde qu'est-ce je fais"

Rule 21.77 GV -> V PR LC PR PF SV 2 types (3 tokens)

Rules 21.76 and 21.77 analyze the same sort of sentences.

"regarde qu'est-ce qu'il fait là"

Rule 21.78 GV -> V PR LC PR PD SV 3 types (5 tokens)

"je veux voir qu'est-ce que c'est ça"

This rule also recognizes the following two sentences which are not correctly analyzed:

"tiens, qu'est-ce que c'est ça?"
"tiens, qu'est-ce que c'est ce truc-là?"

Rule 21.79 GV -> V PR LC PY SV 2 types (3 tokens)

"devine qu'est-ce y a dedans"

Rule 21.80 GV -> V PR LC PR PY SV 1 type (3 tokens)

"devine qu'est-ce qu'y a dedans"

III.22 Group 22 Verb Phrases with Prepositions or Personal Pronouns

As already mentioned, GV' generates verb phrases that begin with a preposition or a personal pronoun.

Rule 22.1 GV' -> PS V 86 types (244 tokens)

This rule recognizes verbs that have a personal pronoun as a direct or indirect object, as well as verbs that are in the pronominal form:

"s'il te plaît"
"je l'envoie"
"je vais le dire"
"je me tortille"
"il se salit"

Rule 22.2 GV' -> PS V SNP 142 types (199 tokens)

This rule recognizes four forms of sentences:

a) sentences in which the personal pronoun is the direct object of the verb, and SNP is a prepositional noun phrase:

"je le mets dans la voiture"

b) sentences in which the personal pronoun is the indirect object of the verb, and SNP the direct object:

"je te fauche la cuillère"

c) sentences in which the verb is in the pronominal form and SNP repeats the subject:

"il se cache le camion"

d) sentences in which the personal pronoun is the direct object of the verb and SNP repeats the object:

"je le mets le disque"

Rule 22.3 GV' -> PS V DG 18 types (34 tokens)

This rule performs the same analysis as rule 22.1 for sentences in the negative form:

"j'en veux pas"
"non il en manque pas"
"on les reconnaît pas"

Rule 22.4 GV' -> PS V DV 28 types (64 tokens)

This rule generates the same forms of sentences as 22.1 when the ending is an adverb:

"il se met là"
"tu m'attends ici"
"l'échelle se tord encore"

In a few cases, this rule gives an incorrect analysis of sentences that contain the verbal element me fait mal which is analyzed, after the disambiguation process as PS VT DV, instead of PS VT NC.

Rule 22.5 GV' -> PS V DG SNP 16 types (18 tokens)

This rule generates two types of sentences:

a)sentences in which SNP is a repetition of the object:

"non j'en veux pas du chapeau"
"je le veux plus ce bol"

b) sentences in which SNP is a prepositional noun phrase:

"oui on la met pas dans ce garage"

Rule 22.6 GV' -> PS V IN 3 types (5 tokens)

"j'en ai une"
"je vais en chercher une"

Rule 22.7 GV' -> PS V AC 5 types (13 tokens)

"en a deux"
"non j'en ai trois"

Rule 22.8 GV' -> PS V IN AQ 1 types (2 tokens)

"en avait un vert"

Rule 22.9 GV' -> PS V DV SNP 23 types (34 tokens)

"en a beaucoup de petits lions"
"je vais la mettre là la remorque"
"et ça s'appelle aussi des sous"
"on le met comme ça"

In a few cases, this rule analyzes incorrectly sentences which contain the word plus:

"en a plus d'école"
"en a plus de roses"

where the plus is recognized as DV and DG, but is analyzed as DV after the disambiguation process.

Rule 22.10 GV' -> PD V DV' SNP 1 type (1 token)

"tu me donnes un peu d'eau"

Rule 22.11 GV' -> PS V DV DV' 4 types (6 tokens)

"j'en veux encore un peu"
"je veux l'écouter encore un peu"

Rule 22.12 GV' -> PS PS V 9 types (14 tokens)

"je vais te le montrer"
"il faut pas m'en parler"

Rule 22.13 GV' -> PS PS V DG 4 types (6 tokens)

"m'en souviens pas"
"je t'en fais pas"

In one case it recognizes an utterance where Philippe stuttered:

"non le l'accroche pas"

Rule 22.14 GV' -> NE PS V DG 2 types (3 tokens)

"ne me demande pas"
"ne l'ouvre pas"

Rule 22.15 GV' -> PS DV V' 12 types (21 tokens)

This rule parses the verb s'en aller which is coded PS,CS PS,DV,EP VI:

"il s'en va"
"non pas t'en aller"
"la petite fille elle s'en va"

Rule 22.16 GV' -> EP V 23 types (44 tokens)

"à boire"
"pour conduire"
"sans savoir"
"des boutons pour conduire"

In a few cases, this rule recognizes compound words:

"des machines à écrire"
"machine à laver"

Rule 22.17 GV' -> EP V SNP 57 types (94 tokens)

"pour lancer la balle"
"le bâton pour attraper des poissons"
"pour faire quoi?"

Rule 22.18 GV' -> EP PS V 10 types (19 tokens)

"à se battre"
"comment on fait pour l'ouvrir?"

Rule 22.19 GV' -> EP PS V SNP 8 types (9 tokens)

"pour se protéger de la pluie"
"comment on fait pour l'ouvrir celui-là?"

In two cases, this rule gives an incorrect analysis:

"sur la figure la poupée"
"du crayon sur la figure la poupée"

The terminal for the above is EP DN,PS NC,VN DN NC, which form is recognized as EP PS VN DN NC rather than EP DN NC DN NC.

Rule 22.20 GV' -> EP V PR SNS 7 types (10 tokens)

"c'est pour faire quoi ces boutons?"
"c'est pour faire quoi le vieux couteau?"

Rule 22.21 GV' -> EP V DV 3 types (4 tokens)

"pour monter dedans"
"oui pour garer là"
"à regarder dehors"

Rule 22.22 GV' -> EP V SNP DD 4 types (6 tokens)

"pour mettre sur la table ça"
"c'est pour faire quoi celle-là?"

Rule 22.23 GV' -> EP DG PS V SNP 1 type (2 tokens)

"pour pas se brûler les doigts"

Rule 22.24 GV' -> EP CS SNS V' DG 0 type

This rule recognizes sentences that begin with the conjunction of subordination pour que. The probabilistic disambiguation selects the form EP PR in preference to EP CS, and consequently this rule has not been used.

Rule 22.25 GV' -> EP CS SNS V' 0 type

For the same reason as in the case of rule 22.24, this rule has not been used.

III.23 Group 23 Intransitive and Impersonal Verbs

Rule 23.1 V' -> VI 236 types (415 tokens)

VI is the terminal symbol for intransitive verbs.

Rule 23.2 V' -> VM 3 types (3 tokens)

VM is the terminal symbol for impersonal verbs.

III.24 Group 24 Copula

Rule 24.1 COP -> VC 661 types (1471 tokens)

As already mentioned, VC is the terminal symbol for être, devenir and rester.

III.25 Group 25 Transitive and Pronominal Verbs

V is rewritten as VT (transitive), VN (transitive or intransitive), VR (pronominal)

Rule 25.1 V -> VT 1391 types (2954 tokens)

Rule 25.2 V -> VR 13 types (43 tokens)

Only the verbs which do not exist under another form have been coded VR.

Rule 25.3 V -> VN 591 types (1142 tokens)

III.26 Group 26 Predicate Adjectives

This group presents the production rules of adjective phrases which are predicate adjectives.

Rule 26.1 SA -> SA' 104 (389)

"c'est mauvais"
"il est pas beau"

Rule 26.2 SA -> DM SA' 25 (5)

This rule recognizes predicate adjectives which are modified by an adverb:

"trop petit"
"c'est très drôle"

III.27 Group 27 Adjectives

SA' is rewritten as an adjective without determiner. There are two "ad hoc" rules: 27.2 and 27.3: AQ EP NC, AQ EP AQ NC: plein de..., plein de petits....

Rule 27.1 SA' -> AQ 126 types (439 tokens)

Rule 27.2 SA' -> AQ EP NC 3 types (7 tokens)

"plein de cheminées"

Rule 27.3 SA' -> AQ EP AQ NC 1 type (2 tokens)

"plein de petits bateaux"

III.28 Group 28 Questions

This group of rules parses sentences in the interrogative form. It generates incomplete questions. For example, the grammar will recognize as well qu'est-ce que? as que ce c'est? or qu'est-ce?. As mentioned already, not all the questions are generated by this group of rules; in particular questions which begin with an interrogative adverb are analyzed in the same way as sentences in the declarative form.

Rule 28.1 Q -> PR 1 type (73 tokens)

PR is the terminal symbol for interrogative pronouns:

"quoi?"
"qui?"
"lequel?"

Rule 28.2 Q -> DET PR 2 types (4 tokens)

"le quoi?"
"mon quoi?"

Rule 28.3 Q -> PR DV 1 type (2 tokens)

"qui encore?"

Rule 28.4 Q -> CC PR DV 1 type (2 tokens)

"et qui encore?"

Rule 28.5 Q -> PD VC PR 1 type (13 tokens)

"c'est quoi?"

Rule 28.6 Q -> DV PD VC PR 1 type (1 token)

"là c'est quoi?"

Rule 28.7 Q -> CC PD PD VC PR 1 type (2 tokens)

"et ça c'est quoi?"

Rule 28.8 Q -> PR LC PR PF SV 25 types (35 tokens)

This rule recognizes interrogations beginning with qu'est-ce que:

"qu'est-ce que tu fais?"
"qu'est-ce qu'il fait là"?

Rule 28.9 Q -> PR LC PF SV 18 types (40 tokens)

This rule generates interrogations in which qu'est-ce is used instead of qu'est-ce que:

"qu'est-ce tu fais?"
"qu'est-ce tu vas m'acheter"

Rule 28.10 Q -> PR LC PD SV 6 types (26 tokens)

This rule generates the same form of questions as rule 28.9:

"qu'est-ce c'est ça?"
"qu'est-ce c'est dedans?"

Rule 28.11 Q -> PD PR LC PD SV 2 types (5 tokens)

"celui-là qu'est-ce c'est?"

Rule 28.12 Q -> PR LC PY SV 10 types (21 tokens)

This rule generates interrogatives where qu'est-ce y a is said instead of qu'est-ce qu'il y a:

"qu'est-ce y a encore?"
"qu'est-ce y a de beau Clara?"

Rule 28.13 Q -> CC PD PR LC PD SV 1 type (5 tokens)

"et ça qu'est-ce c'est?"

Rule 28.14 Q -> PR LC PR SV 13 types (17 tokens)

"qu'est-ce qui se passe?"
"qu'est-ce qui fait ça?"

Rule 28.15 Q -> PR PD SV 10 types (61 tokens)

"que c'est ça?"
"qui c'est ça?"

Rule 28.16 Q -> PR PD COP PR SV 9 types (9 tokens)

"qui c'est qui l'a fait ça?"
"qui c'est qui téléphone?"

Rule 28.17 Q -> CC PD PR PD SV 1 type (2 tokens)

"et ça que c'est?"

Rule 28.18 Q -> PD VC PR PR SV 5 types (5 tokens)
"c'est qui qui me donne cette voiture-là?"
"c'est quoi qui marche pas?"

Rule 28.19 Q -> PD VC DV DV 1 types (8 tokens)

"c'est où là?"

Rule 28.20 Q -> PD VC PR SN 20 types (104 tokens)

"c'est quoi ça?"
"c'est quoi la police?"

Rule 28.21 Q -> CC PD PD VC PR SN 1 type (1 token)

"et ça c'est quoi ça?"

Rule 28.22 Q -> PD VC PR SN DV 7 types (15 tokens)

"c'est quoi cette goutte là?"
"c'est quoi les fenêtres là?"

Rule 28.23 Q -> CC PD PD VC PR SN DV 1 type (1 token)

"et ça c'est quoi ces boules là?"

Rule 28.24 Q -> PR PY SV 3 types (6 tokens)

"qui y a dehors?"
"qui y avait dans la voiture?"

III.29 Group 29 Adverbs of Quantity

This group presents the production rules of two expressions: un peu and un petit peu, these expressions function as adverbs of quantity.

Rule 29.1 DV' -> IN DM 19 types (26 tokens)

"je veux que tu restes encore un peu"

Rule 29.2 DV' -> IN AQ DM 10 types (14 tokens)

"un petit peu"
"moi veux un petit peu de lumière"

III.30 Summary results of the grammatical analysis

There are 15057 tokens (6539 types) in the corpus; 11294 tokens (4000 types) have been recognized by the grammar, which corresponds to 75.01 percent of the tokens and to 61.17 percent of the types. We should mention that a large number of types, 571 accounting for 1480 tokens, include one or more occurrences of MS (uncoded sounds). Since the grammar makes no effort to account for the MS, we should note that 67.02 percent of the types and 83.18 percent of the tokens were recognized when utterances involving MS were discarded.

The results of the grammatical analysis are presented in Table IV.

TABLE IV

Description	Types	Tokens
Total in corpus	6539	15057
Total recognized	4000	11294
Percent recognized	61.17	75.01
Ambiguous in corpus	5703	8774
Percent ambiguous in corpus	87.22	58.27
Ambiguous recognized	3556	6435
Percent ambiguous recognized	62.35	73.34
Ambiguity resolved	2808	5309
Ambiguity reduced	660	938

Original ambiguity factor = 62033

Parsed ambiguity factor = 31715
Reduced ambiguity factor = 1402

The entries in Table IV are defined as follows, with the usual distinctions between types and tokens. We first give the total number of types and tokens in the corpus, we second give the number recognized, and third the percent recognized. The fourth line indicates how many of the types or tokens in the corpus are lexically ambiguous, that is, have at least one word assigned to more than a grammatical category. For instance, the type

(1) EP DN,PS NC

occurs 119 times in the corpus. The second word, according to the dictionary, could be either DN or PS ("le", "la"). Hence (1) expands to

(1') EP DN NC
(1") EP PS NC.

While the grammar recognizes EP DN NC, it does not recognize EP PS NC. In this case, we say that the lexical ambiguity is <u>resolved</u>, meaning that only one of the alternative types has been recognized.

Another lexically ambiguous type is:

(2) PR,DV,CS PD VC,VA PD

which occurred 33 times. It expands to:

(2.1) PR PD VC PD
(2.2) DV PD VC PD
(2.3) CS PD VC PD
(2.4) PR PD VA PD
(2.5) DV PD VA PD
(2.6) CS PD VA PD

In this case, (2.1), (2.2) and (2.3) are recognized, and so we say that the lexical ambiguity was reduced (but not resolved). This reduction was effective for 660 types (938 tokens) in the corpus; 748 types (1126 tokens) remained lexically ambiguous and were not resolved by the grammar. The probabilistic method for removing this lexical ambiguity is presented in the next chapter.

In Appendix I we give the main examples of utterances for which the grammatical analysis is not satisfactory.

Chapter IV

PROBABILISTIC GRAMMAR AND PROBABILISTIC DISAMBIGUATION

The relatively detailed and complex grammar described in the preceding chapter parses about 75 percent of the utterances in Philippe's corpus. This criterion alone is not sufficient to judge the grammar, for it would be easy enough to write a grammar that would parse 100 percent of the utterances, namely the universal grammar. There are different ways of thinking about how additional criteria may be imposed on a grammar in order to determine its appropriateness for a given corpus. For a number of reasons, we consider a probabilistic criterion of goodness of fit one of the better ways to evaluate the appropriateness of a grammar. This probabilistic viewpoint has been developed extensively in previous publications originating as technical reports of the Institute of Mathematical Studies in the Social Sciences [8], [2], [7], [9]. We shall not attempt to review in detail the many different grounds on which such a probabilistic viewpoint can be supported.

The basic strategy of the probabilistic approach to grammars is to attach a parameter to each rule of a group of rules with the requirement that the parameters be interpreted as probabilities, that is, each parameter is non-negative and the sum of the parameters for a given group of rules is equal to 1. The parameter is meant to correspond to the frequency of use of the given production rule of the

grammar in generating the utterances of the corpus. Once such parameters are assigned, we can estimate them by standard methods, for example, in many straightforward cases by maximum-likelihood methods. Having estimated the parameters we are then able to move on to consider a standard goodness-of-fit criterion for evaluating the adequacy of the grammar to the corpus. We should say at once that in the present stage of investigation the goodness-of-fit criterion is not well satisfied; that is, if we take a large corpus, for instance, that of Philippe's speech, we do not anticipate obtaining a reasonable level of significance for the fit of the probabilistic grammar to the corpus. We can, as was done in the case of our earlier report [9], use the goodness-of-fit criterion to distinguish between two grammars. In the present case we want to use the probabilistic apparatus to disambiguate grammatically ambiguous utterances. As we describe below, we believe that this represents a useful application of probabilistic grammars and one that has some theoretical interest.

Before turning to the details of the probabilistic grammar, we want to reiterate a remark that has been made in previous publications on these matters. It is that the probabilities themselves are not to be interpreted as ultimate parameters of Philippe's speech. The parameters represent the results of averaging over a number of utterances. They do not give an account of the perceptual and semantical mechanisms for generating a particular utterance on a particular occasion. They play the averaging role that is

characteristic of such analyses in many other parts of science and should be so interpreted.

Table V shows the usage of each rule in the grammar. For example, rule 1.1 is used 2611 times, and the cumulative usage of the rules of group 1 is 11294. Here, "usage" means the frequency of a given lexical type multiplied by the number of times the rule occurred within the derivation of that type. Table V also shows the estimated probability of usage for each rule. These probabilities are used in the method of lexical disambiguation described below.

Secondly, in Table VI we show the observed and expected frequencies of sentence types for those that have a frequency of at least 10 in the corpus. It will be seen immediately from a perusal of this table that the fit of the probabilistically computed expectations is not exact to the observed data, and, in fact, is not as good as one would like in a completely satisfactory theory. On the other hand, it is our judgment that, without increasing the number of rules extravagantly, it would be hard to improve substantially the fit as indicated in this table. Because we are more interested in the present report in the developmental analysis contained in the next chapter, we shall not examine in more detail at this point the fit of the probabilistic grammar to the entire corpus.

We turn now to discussion of probabilistic disambiguation. When an utterance has more than one dictionary representation, the utterance is <u>lexically ambiguous</u>. If only one of those dictionary

representations is parsed by the grammar, we say that the ambiguity is only apparent. However, if more than one representation is recognized, then we have to account for that ambiguity.

One tenable view is that the several lexical ambiguities are all intuitively reasonable. While this is possible, it is nevertheless plausible that Philippe only acts upon one interpretation -- he makes some decision about which interpretation to accept.

We have proposed in [7] and [9] that lexical ambiguity be treated syntactically and probabilistically. In particular, see [7] for a discussion of the details of this method. Of the several lexical interpretations for a sentence, we accept the most likely interpretation according to the probabilistic grammar that we have offered. In doing so, we are not claiming that disambiguation does not involve semantic considerations in a crucial way. Rather, we are claiming that syntactic features (of which the probabilistic grammar is a key example) may well play a role in disambiguation. This could happen in several ways, but the way we consider to be the most reasonable would involve interaction between probabilistic analysis and semantic and contextual analysis, where the initial decision on what to consider semantically is made by the probabilistic grammar. We should remark that this interpretation is, of course, a listener-oriented view.

The surprising feature about probabilistic lexical disambiguation is the degree to which it appears to work in a plausible

way. There were 660 lexically ambiguous types (938 tokens) in which the ambiguity was not apparent (i.e., there were two or more lexical alternatives recognized by the grammar). Of these, only 88 types (133 tokens) were resolved in an unsatisfactory way. This corresponds to 13.33 percent of the types and 14.17 per cent of the tokens.

Table VII presents the types obtained after ambiguity resolution as defined in the previous chapter and before probabilistic disambiguation, down to frequency 5. The first digit on the left gives the frequency of a type, followed by the ambiguous type and then by the type obtained after disambiguation.

We have also been concerned to analyze the intuitively incorrect decisions made by probabilistic lexical disambiguation. These apparent errors fall into several simple categories.

Two criteria have guided us in deciding whether or not the ambiguity is solved correctly by the method described before: a reference to the context in which the utterance has been emitted, and the grammatical analysis of the elements of the utterance. Two examples will illustrate the point.

The string PS,DN VN,PO which corresponds to the utterance la tienne is recognized as PS VN (personal pronoun followed by verb) and as DN PO (possessive pronoun). Reference to the context shows that Philippe was using the personal pronoun. Consequently, our conclusion is that the ambiguity is incorrectly solved, since the string PS VN has a higher probability than the sring DN PO.

As a second example, the utterance il est vide le pot which is coded PF VC,VA AQ,VT DN NC is recognized as PF VC AQ DN NC as well as PF VA VT DN NC. After the disambiguation process, PF VA VT DN NC is chosen, and we say that the grammatical analysis is incorrect because in this utterance vide is not a verb, but an adjective.

In Apprendix II, we present the utterances for which the probabilistic disambiguation is not correct.

TABLE V

Usage for Each Rule

Label	Usage	Cum. Usage	Est. Prob.	Rule
(1,1)	2611	11294	.23	s -> s1
(1,2)	2669	11294	.24	s -> s2
(1,3)	6014	11294	.53	s -> s3
(2,1)	1494	2611	.57	s1 -> dv
(2,2)	3	2611	.11@-2	s1 -> dv'
(2,3)	705	2611	.27	s1 -> dg
(2,4)	198	2611	.76@-1	s1 -> ko
(2,5)	35	2611	.13@-1	s1 -> lc
(2,6)	27	2611	.10@-1	s1 -> num
(2,7)	24	2611	.92@-2	s1 -> cc dv
(2,8)	19	2611	.73@-2	s1 -> cc dg
(2,9)	9	2611	.34@-2	s1 -> ko dg
(2,10)	42	2611	.16@-1	s1 -> ko dv
(2,11)	21	2611	.80@-2	s1 -> dv dv
(2,12)	5	2611	.19@-2	s1 -> dg dv
(2,13)	5	2611	.19@-2	s1 -> dg lc
(2,14)	11	2611	.42@-2	s1 -> dg dg
(2,15)	0	2611	.00	s1 -> dm dv
(2,16)	9	2611	.34@-2	s1 -> ep dv
(2,17)	4	2611	.15@-2	s1 -> ps dv
(3,1)	2025	2669	.76	s2 -> sn*1
(3,2)	1	2669	.37@-3	s2 -> dv' sn
(3,3)	100	2669	.37@-1	s2 -> sn dv
(3,4)	54	2669	.20@-1	s2 -> sn np
(3,5)	162	2669	.61@-1	s2 -> dv sn*1
(3,6)	63	2669	.24@-1	s2 -> dg sn*1
(3,7)	41	2669	.15@-1	s2 -> ko sn*1
(3,8)	13	2669	.49@-2	s2 -> lc sn*1
(3,9)	52	2669	.19@-1	s2 -> cc sn*1
(3,10)	11	2669	.41@-2	s2 -> dg dg sn*1
(3,11)	9	2669	.34@-2	s2 -> cc dv sn*1
(3,12)	4	2669	.15@-2	s2 -> cc dg sn*1
(3,13)	9	2669	.34@-2	s2 -> dg sn*1 dv
(3,14)	9	2669	.34@-2	s2 -> cc sn*1 dv
(3,15)	62	2669	.23@-1	s2 -> sa
(3,16)	23	2669	.86@-2	s2 -> det sa
(3,17)	10	2669	.37@-2	s2 -> det sa cc det sa
(3,18)	9	2669	.34@-2	s2 -> dg sa
(3,19)	2	2669	.75@-3	s2 -> cc sa
(3,20)	10	2669	.37@-2	s2 -> dv sa
(4,1)	4385	6014	.73	s3 -> s'
(4,2)	176	6014	.29@-1	s3 -> dg s'

(4,3)	571	6014	.95@-1	s3 -> dv s'
(4,4)	10	6014	.17@-2	s3 -> lc s'
(4,5)	33	6014	.55@-2	s3 -> cc s'
(4,6)	146	6014	.24@-1	s3 -> ko s'
(4,7)	15	6014	.25@-2	s3 -> cs s'
(4,8)	16	6014	.27@-2	s3 -> s' ko
(4,9)	55	6014	.91@-2	s3 -> lc cs s'
(4,10)	60	6014	.10@-1	s3 -> cc dv s'
(4,11)	11	6014	.18@-2	s3 -> dg dg s'
(4,12)	20	6014	.33@-2	s3 -> ko dv s'
(4,13)	12	6014	.20@-2	s3 -> cc dg s'
(4,14)	16	6014	.27@-2	s3 -> dv dv s'
(4,15)	30	6014	.50@-2	s3 -> s' dv dv
(4,16)	4	6014	.67@-3	s3 -> cc dv dv s'
(4,17)	454	6014	.75@-1	s3 -> q
(5,1)	1643	5592	.29	s' -> sv
(5,2)	3662	5592	.65	s' -> sns sv
(5,3)	82	5592	.15@-1	s' -> sns sv cs sns sv
(5,4)	15	5592	.27@-2	s' -> sns sv cc sns sv
(5,5)	75	5592	.13@-1	s' -> sns sv pr sv
(5,6)	35	5592	.63@-2	s' -> sns pr sv
(5,7)	66	5592	.12@-1	s' -> sn** sns sv
(5,8)	14	5592	.25@-2	s' -> sn** gv'
(6,1)	892	1318	.68	sn -> sn*
(6,2)	247	1318	.19	sn -> sn**
(6,3)	31	1318	.24@-1	sn -> pt
(6,4)	148	1318	.11	sn -> pd
(7,1)	801	4439	.18	sns -> sn*
(7,2)	2284	4439	.51	sns -> pf
(7,3)	162	4439	.36@-1	sns -> sn* pf
(7,4)	55	4439	.12@-1	sns -> pf pf
(7,5)	956	4439	.22	sns -> pd
(7,6)	141	4439	.32@-1	sns -> pd pd
(7,7)	22	4439	.50@-2	sns -> sn* pd
(7,8)	6	4439	.14@-2	sns -> pd dv pd
(7,9)	12	4439	.27@-2	sns -> sn* dv pf
(8,1)	1602	2505	.64	snp -> sn*
(8,2)	713	2505	.28	snp -> sn**
(8,3)	17	2505	.68@-2	snp -> pr
(8,4)	173	2505	.69@-1	snp -> pd
(9,1)	3223	3491	.92	sn* -> np*
(9,2)	13	3491	.37@-2	sn* -> np* cc np*
(9,3)	230	3491	.66@-1	sn* -> np* ep np*
(9,4)	1	3491	.29@-3	sn* -> np* cc np* ep np*
(9,5)	8	3491	.23@-2	sn* -> np* ep ep np*
(9,6)	11	3491	.32@-2	sn* -> np* ep pt
(9,7)	5	3491	.14@-2	sn* -> np* ep pr
(10,1)	1083	1277	.85	sn** -> ep np*
(10,2)	91	1277	.71@-1	sn** -> ep np* ep np*
(10,3)	4	1277	.31@-2	sn** -> ep nz* cc np*

(10,4)	5	1277	.39@-2	sn** -> ep np* cc ep np*
(10,5)	29	1277	.23@-1	sn** -> ep ep np*
(10,6)	27	1277	.21@-1	sn** -> ep pt
(10,7)	39	1277	.31@-1	sn** -> ep pr
(11,1)	1	7494	.13@-3	np* -> num adjp nc postadjp
(11,2)	14	7494	.19@-2	np* -> det adjp nc postadjp
(11,3)	1	7494	.13@-3	np* -> adjp nc postadjp
(11,4)	2	7494	.27@-3	np* -> det num nc postadjp
(11,5)	2	7494	.27@-3	np* -> num nc postadjp
(11,6)	88	7494	.12@-1	np* -> det nc postadjp
(11,7)	29	7494	.39@-2	np* -> nc postadjp
(11,8)	1	7494	.13@-3	np* -> det num adjp nc
(11,9)	7	7494	.93@-3	np* -> num adjp nc
(11,10)	365	7494	.49@-1	np* -> det adjp nc
(11,11)	128	7494	.17@-1	np* -> adjp nc
(11,12)	7	7494	.93@-3	np* -> det num nc
(11,13)	55	7494	.73@-2	np* -> num nc
(11,14)	3602	7494	.48	np* -> det nc
(11,15)	2899	7494	.39	np* -> np"
(11,16)	73	7494	.97@-2	np* -> det ai nc
(11,17)	38	7494	.51@-2	np* -> ai det nc
(11,18)	4	7494	.53@-3	np* -> ai det nc ad
(11,19)	37	7494	.49@-2	np* -> ai nc
(11,20)	14	7494	.19@-2	np* -> det num
(11,21)	34	7494	.45@-2	np* -> det np
(11,22)	4	7494	.53@-3	np* -> det adjp np
(11,23)	16	7494	.21@-2	np* -> adjp np
(11,24)	70	7494	.93@-2	np* -> det nc ad
(11,25)	6	7494	.80@-3	np* -> det adjp nc ad
(12,1)	1385	2398	.58	sn*1 -> np*
(12,2)	17	2398	.71@-2	sn*1 -> np* cc np*
(12,3)	161	2398	.67@-1	sn*1 -> np* ep np*
(12,4)	8	2398	.33@-2	sn*1 -> np* ep ep np*
(12,5)	2	2398	.83@-3	sn*1 -> np* cc ep np*
(12,6)	6	2398	.25@-2	sn*1 -> np* ep pr
(12,7)	8	2398	.33@-2	sn*1 -> np* ep pt
(12,8)	19	2398	.79@-2	sn*1 -> pf
(12,9)	145	2398	.60@-1	sn*1 -> pd
(12,10)	8	2398	.33@-2	sn*1 -> pd ep np*
(12,11)	521	2398	.22	sn*1 -> ep np*
(12,12)	44	2398	.18@-1	sn*1 -> ep np* ep np*
(12,13)	6	2398	.25@-2	sn*1 -> ep np* cc np*
(12,14)	4	2398	.17@-2	sn*1 -> ep np* cc ep np*
(12,15)	4	2398	.17@-2	sn*1 -> ep np* ep ep np*
(12,16)	19	2398	.79@-2	sn*1 -> ep ep np*
(12,17)	18	2398	.75@-2	sn*1 -> ep pt
(12,18)	15	2398	.63@-2	sn*1 -> ep pr
(12,19)	8	2398	.33@-2	sn*1 -> ep pd
(13,1)	17	2899	.59@-2	np" -> dn po
(13,2)	383	2899	.13	np" -> pi

(13,3)	71	2899	.24@-1	np" -> det pl
(13,4)	874	2899	.30	np" -> np
(13,5)	1554	2899	.54	np" -> [illegible]
(14,1)	542	552	.98	adjp -> aq
(14,2)	10	552	.18@-1	adjp -> adjp aq
(15,1)	137	137	1.0	postadjp -> aq
(16,1)	237	4447	.53@-1	det -> ad
(16,2)	32	4447	.72@-2	det -> at
(16,3)	342	4447	.77@-1	det -> ac
(16,4)	2599	4447	.58	det -> dn
(16,5)	1237	4447	.28	det -> in
(17,1)	115	120	.96	num -> ac
(17,2)	5	120	.42@-1	num -> an
(18,1)	4124	6025	.68	sv -> gv
(18,2)	589	6025	.98@-1	sv -> gv'
(18,3)	478	6025	.79@-1	sv -> aux1 gv
(18,4)	10	6025	.17@-2	sv -> aux1 dv gv
(18,5)	33	6025	.55@-2	sv -> aux1 dg gv
(18,6)	167	6025	.28@-1	sv -> aux2 gv'
(18,7)	33	6025	.55@-2	sv -> aux2 dg gv'
(18,8)	384	6025	.64@-1	sv -> aux2 gv
(18,9)	6	6025	.10@-2	sv -> aux2 dv gv
(18,10)	43	6025	.71@-2	sv -> aux2 dg gv
(18,11)	118	6025	.20@-1	sv -> ps aux1 gv
(18,12)	16	6025	.27@-2	sv -> ps aux1 dg gv
(18,13)	6	6025	.10@-2	sv -> ps aux2 gv
(18,14)	5	6025	.83@-3	sv -> ps ps aux1 gv
(18,15)	13	6025	.22@-2	sv -> py aux1 gv
(19,1)	662	673	.98	aux1 -> va
(19,2)	11	673	.16@-1	aux1 -> va vd
(20,1)	619	639	.97	aux2 -> vd
(20,2)	12	639	.19@-1	aux2 -> vd vd
(20,3)	8	639	.13@-1	aux2 -> vd ps vd
(21,1)	72	5236	.14@-1	gv -> cop
(21,2)	8	5236	.15@-2	gv -> cop lc
(21,3)	194	5236	.37@-1	gv -> cop sa
(21,4)	663	5236	.13	gv -> cop sn
(21,5)	59	5236	.11@-1	gv -> cop sa sn
(21,6)	3	5236	.57@-3	gv -> cop sa dv
(21,7)	29	5236	.55@-2	gv -> cop sn dv
(21,8)	43	5236	.82@-2	gv -> cop dv sn
(21,9)	47	5236	.90@-2	gv -> cop dg sa
(21,10)	45	5236	.86@-2	gv -> cop dg sn
(21,11)	52	5236	.99@-2	gv -> cop dv
(21,12)	8	5236	.15@-2	gv -> cop dg dv
(21,13)	28	5236	.53@-2	gv -> cop gv'
(21,14)	20	5236	.38@-2	gv -> cop dg sa sn
(21,15)	132	5236	.25@-1	gv -> cop sn sns
(21,16)	5	5236	.95@-3	gv -> cop dv sn dv
(21,17)	3	5236	.57@-3	gv -> cop dg sn gv'

(21,18)	4	5236	.76@-3	gv -> cop dv' aq
(21,19)	9	5236	.17@-2	gv -> cop dg dv sn
(21,20)	21	5236	.40@-2	gv -> cop sn gv'
(21,21)	4	5236	.76@-3	gv -> cop dm v
(21,22)	2	5236	.38@-3	gv -> cop dm v sn
(21,23)	16	5236	.31@-2	gv -> py cop
(21,24)	759	5236	.14	gv -> v
(21,25)	1360	5236	.26	gv -> v snp
(21,26)	136	5236	.26@-1	gv -> v snp dv
(21,27)	91	5236	.17@-1	gv -> v dv
(21,28)	63	5236	.12@-1	gv -> v dv snp
(21,29)	93	5236	.18@-1	gv -> v dg snp
(21,30)	9	5236	.17@-2	gv -> v dg snp sn**
(21,31)	1	5236	.19@-3	gv -> ne v dg snp
(21,32)	9	5236	.17@-2	gv -> v lc
(21,33)	36	5236	.69@-2	gv -> v dg dv
(21,34)	186	5236	.36@-1	gv -> v snp sns
(21,35)	123	5236	.23@-1	gv -> v snp sn**
(21,36)	178	5236	.34@-1	gv -> v dg
(21,37)	29	5236	.55@-2	gv -> v pt
(21,38)	3	5236	.57@-3	gv -> v pt v
(21,39)	17	5236	.32@-2	gv -> v pt snp
(21,40)	3	5236	.57@-3	gv -> v pt pt
(21,41)	4	5236	.76@-3	gv -> v pt dv
(21,42)	19	5236	.36@-2	gv -> v pt dv snp
(21,43)	3	5236	.57@-3	gv -> v dv'
(21,44)	2	5236	.38@-3	gv -> v dv dv'
(21,45)	11	5236	.21@-2	gv -> v dv' sn
(21,46)	1	5236	.19@-3	gv -> v dv dv' sn
(21,47)	2	5236	.38@-3	gv -> v dm dv
(21,48)	20	5236	.38@-2	gv -> v dm aq
(21,49)	4	5236	.76@-3	gv -> v num
(21,50)	8	5236	.15@-2	gv -> v det aq
(21,51)	7	5236	.13@-2	gv -> v ep aq
(21,52)	32	5236	.61@-2	gv -> v snp pr s'
(21,53)	158	5236	.30@-1	gv -> v'
(21,54)	67	5236	.13@-1	gv -> v' sns
(21,55)	22	5236	.42@-2	gv -> v' dv
(21,56)	105	5236	.20@-1	gv -> v' sn**
(21,57)	22	5236	.42@-2	gv -> v' dg
(21,58)	11	5236	.21@-2	gv -> v' dg sns
(21,59)	3	5236	.57@-3	gv -> v' dv'
(21,60)	1	5236	.19@-3	gv -> v' dv' dv
(21,61)	4	5236	.76@-3	gv -> v' dv dv'
(21,62)	5	5236	.95@-3	gv -> v' dm aq
(21,63)	[illegible]	5236	.13@-2	gv -> py v
(21,64)	5	5236	.11@-1	gv -> py v snp
(21,65)	5	5236	.95@-3	gv -> py v dv
(21,66)	5	5236	.95@-3	gv -> py v lc
(21,67)	11	5236	.21@-2	gv -> py v snp pr sv

(21,68)	1	5236	.19@-3	gv -> py v dg
(21,69)	7	5236	.13@-2	gv -> py v dg snp
(21,70)	9	5236	.17@-2	gv -> py v dv snp
(21,71)	18	5236	.34@-2	gv -> py v snp dv
(21,72)	8	5236	.15@-2	gv -> py ps v dv
(21,73)	3	5236	.57@-3	gv -> py ps v dv snp
(21,74)	1	5236	.19@-3	gv -> py ps v dg
(21,75)	1	5236	.19@-3	gv -> py ps v dg snp
(21,76)	18	5236	.34@-2	gv -> v pr lc pf sv
(21,77)	3	5236	.57@-3	gv -> v pr lc pr pf sv
(21,78)	5	5236	.95@-3	gv -> v pr lc pr pd sv
(21,79)	3	5236	.57@-3	gv -> v pr lc py sv
(21,80)	3	5236	.57@-3	gv -> v pr lc pr py sv
(22,1)	248	855	.29	gv' -> ps v
(22,2)	199	855	.23	gv' -> ps v snp
(22,3)	34	855	.40@-1	gv' -> ps v dg
(22,4)	64	855	.75@-1	gv' -> ps v dv
(22,5)	18	855	.21@-1	gv' -> ps v dg snp
(22,6)	5	855	.58@-2	gv' -> ps v in
(22,7)	13	855	.15@-1	gv' -> ps v ac
(22,8)	2	855	.23@-2	gv' -> ps v in aq
(22,9)	34	855	.40@-1	gv' -> ps v dv snp
(22,10)	1	855	.12@-2	gv' -> ps v dv' snp
(22,11)	6	855	.70@-2	gv' -> ps v dv dv'
(22,12)	14	855	.16@-1	gv' -> ps ps v
(22,13)	6	855	.70@-2	gv' -> ps ps v dg
(22,14)	3	855	.35@-2	gv' -> ne ps v dg
(22,15)	21	855	.25@-1	gv' -> ps dv v'
(22,16)	44	855	.51@-1	gv' -> ep v
(22,17)	94	855	.11	gv' -> ep v snp
(22,18)	19	855	.22@-1	gv' -> ep ps v
(22,19)	9	855	.11@-1	gv' -> ep ps v snp
(22,20)	11	855	.13@-1	gv' -> ep v pr sns
(22,21)	4	855	.47@-2	gv' -> ep v dv
(22,22)	5	855	.58@-2	gv' -> ep v snp pd
(22,23)	2	855	.23@-2	gv' -> ep dg ps v snp
(22,24)	0	855	.00	gv' -> ep cs sns v' dg
(22,25)	0	855	.00	gv' -> ep cs sns v'
(23,1)	416	419	.99	v' -> vi
(23,2)	3	419	.72@-2	v' -> vm
(24,1)	1475	1475	1.0	cop -> vc
(25,1)	3018	4215	.72	v -> vt
(25,2)	43	4215	.10@-1	v -> vr
(25,3)	1154	4215	.27	v -> vn
(26,1)	391	449	.87	sa -> sa'
(26,2)	58	449	.13	sa -> dm sa'
(27,1)	440	449	.98	sa' -> aq
(27,2)	7	449	.16@-1	sa' -> aq ep nc
(27,3)	2	449	.45@-2	sa' -> aq ep aq nc
(28,1)	73	454	.16	q -> pr

(28,2)	4	454	.88@-2	q -> det pr
(28,3)	2	454	.44@-2	q -> pr dv
(28,4)	2	454	.44@-2	q -> cc pr dv
(28,5)	13	454	.29@-1	q -> pd vc pr
(28,6)	1	454	.22@-2	q -> dv pd vc pr
(28,7)	2	454	.44@-2	q -> cc pd pd vc pr
(28,8)	35	454	.77@-1	q -> pr lc pr pf sv
(28,9)	40	454	.88@-1	q -> pr lc pf sv
(28,10)	26	454	.57@-1	q -> pr lc pd sv
(28,11)	5	454	.11@-1	q -> pd pr lc pd sv
(28,12)	21	454	.46@-1	q -> pr lc py sv
(28,13)	5	454	.11@-1	q -> cc pd pr lc pd sv
(28,14)	17	454	.37@-1	q -> pr lc pr sv
(28,15)	61	454	.13	q -> pr pd sv
(28,16)	9	454	.20@-1	q -> pr pd cop pr sv
(28,17)	2	454	.44@-2	q -> cc pd pr pd sv
(28,18)	5	454	.11@-1	q -> pd vc pr pr sv
(28,19)	4	454	.88@-2	q -> pd vc dv dv
(28,20)	104	454	.23	q -> pd vc pr sn
(28,21)	1	454	.22@-2	q -> cc pd pd vc pr sn
(28,22)	15	454	.33@-1	q -> pd vc pr sn dv
(28,23)	1	454	.22@-2	q -> cc pd pd vc pr sn dv
(28,24)	6	454	.13@-1	q -> pr py sv
(29,1)	26	40	.65	dv' -> in dm
(29,2)	14	40	.35	dv' -> in aq dm
			.00	

TABLE VI

Observed and Expected Frequencies

of Utterances in PHILIPPE

Observed	Expected	Utterance Type
1494	1494.00	dv
705	705.00	dg
295	156.36	in nc
253	328.51	dn nc
246	242.43	nc
198	198.00	ko
154	91.20	ep nc
137	136.39	np
132	123.58	ep dn nc
109	91.52	vt
100	24.62	pf vt dg
87	122.45	pd
76	27.21	vc dn nc
73	73.00	pr
57	13.39	pd vc aq
56	51.31	ep np
48	34.99	vn
44	4.48	pd vc in nc
43	19.61	aq nc
42	52.96	aq
42	42.00	ko dv
40	15.54	in aq nc
40	3.90	pd vc np
40	.59	pr pd vc pd
38	2.93	dv pf vc dn nc
38	32.01	pf vc aq
36	10.40	vn dn nc
35	35.00	lc
33	4.58	pf va vn
31	8.35	in nc dv
30	58.82	ep in nc
30	9.40	pd vc dn nc
30	31.21	pf vt dn nc
28	8.21	vt ep nc
28	12.95	vt in nc
27	25.88	ac
27	32.65	dn aq nc
27	11.68	pd vc pr pd
27	30.00	pf ps vt

26	3.14	cc pd
26	40.14	pf vn
25	8.24	pf vd ps vt
24	24.00	cc dv
23	7.30	dv ep nc
23	1.76	pd vc ep nc
23	4.98	pf vt dg dv
21	21.00	dv dv
21	3.77	in nc ep nc
21	4.88	pf vn ep dn nc
21	14.85	pf vt in nc
21	11.33	vt pd
20	26.28	dv dn nc
20	9.80	dv pd
20	104.98	pf vt
19	19.00	cc dg
19	1.66	pd vc pr ad nc
19	11.97	pf va vt
19	.01	vr pt dv dn nc
18	17.55	dn nc dv
18	9.89	dv ep dn nc
18	26.42	vi
18	11.30	vt np
17	1.81	dv pf vc
17	9.41	pf vt ep nc
16	4.45	dn nc ep np
16	.39	dv pf vc ao nc
16	15.20	ep pt
16	1.64	ep vt dn nc
16	3.17	in ai nc
16	.18	pf va vn ao nc
16	10.04	pf vc dv
16	.66	pi vn
16	4.32	vn np
15	12.67	ep pr
15	.66	pd pd vc in nc
15	3.24	pd vc dg aq
15	12.99	pf vt pd
15	11.12	vt ep dn nc
14	12.68	pd vi
14	11.47	pf ps vn
14	1.36	pf va vn dn nc
14	2.81	pf vd vt dn nc
14	.25	pr lc pd vc pd
14	.33	py vt ep nc
14	3.58	vt ao nc
14	3.50	vt pt
13	7.87	dm aq
13	3.95	dv pf vi
13	3.08	in pi

13	1.99	pd vc dm aq
13	13.00	pd vc pr
13	18.25	pd vc pr dn nc
13	1.26	pd vc pt
13	7.74	pf ps vt dv
13	.64	pf vt dg ep nc
13	.56	vt dg ep nc
13	2.70	vt dn aq nc
12	43.23	ao nc
12	4.04	at nc
12	.10	cs pf ps vt
12	7.92	dn nc ep nc
12	19.39	dv nc
12	5.47	in aq
12	3.82	in nc aq
12	6.65	ko dn nc
12	5.84	nc ep nc
12	1.91	pd va vn
12	.03	pd vc pr ad nc ad
12	3.15	pf vd ps vn
12	6.75	ps vt dv
12	.07	vt dn nc np
12	.03	vt pd np
11	11.00	dg dg
11	12.51	dv in nc
11	16.26	ep ao nc
11	3.33	ep ep nc
11	3.29	ep nc dv
11	4.51	in nc np
11	4.45	nc aq
11	6.99	nc np
11	1.29	pd py vc
11	.19	pd vc dn pi
11	3.45	pf va vi
11	2.84	pr lc pf vt
10	8.03	dn nc aq
10	10.73	dn nc ep dn nc
10	7.92	nc ep dn nc
10	16.04	pf
10	.47	pf va vt ao nc
10	1.46	pf vc dv dn nc
10	30.30	pf vi
10	.82	vt ep nc dv

TABLE VII

Table VII presents the types obtained after consolidation and before disambiguation down to frequency 5. The first digit on the left gives the frequency of a type, followed by the ambiguous type and by the type obtained after disambiguation.

33 PR,DV,CS PD VC PD	PR PD VC PD	
	que c'est ça?	
27 AQ,NC	NC	
	bleu	
	menteur	
	carré	
19 VR PT DV,EP DN NC	VR PT DV DN NC	
18 VI PT,VD PT,VR PT DV,EP DN NC,VT	va-t'en la fumée	
1 VI PT,VD PT,VR PT DV,EP DN NC	va-t'en la machine	
14 CC,DV DV	CC DV	
10 CC,DV CS,DV	mais si	
4 CC,DV DV	mais oui	
14 DN,PS NC,VT	DN NC	
	le peigne	
	la brosse	
	le sucre	
14 PR,DV	DV	
	où?	
10 NC,VT	NC	
	brosse	
	peigne	
	coupe	
10 PF VC AQ,NC	PF VC AQ	
	il est froid	
	il est idiot	
	il est rond	

```
10 PF,PS VA VN AO NC      PF VA VN AO NC
                          t'as fini ton café?
                          t'as fini ton chocolat?

 9 DN AQ,NC               DN NC
                          le con
                          le carré
                          le creux

 9 EP NC,VT               EP NC
                          du bois
                          des barres
                          des bandes

 8 VT DV,DG EP NC         VT DV EP NC
       5 VT,VA DV,DG,DM EP NC  a plus d'eau
                               a plus d'école
                               a plus d'enfants
       1 VT,VA DV,DG,DM EP AQ,NC a plus de roses
       1 VT,VA DV,DG,DM EP NC,VN a plus de fumée
       1 VT,VA DV,DG,DM EP NC    veux plus du chocolat

6 CC,DV DV PF VC          CC DV PF VC
                          mais où il est?

6 IN AQ,NC                IN NC
                          un carré
                          un bleu
                          un élastique

6 PF,PS VT                PF VT
       5 PF,PS VT         t'accroches
       1 PF,PS VT,VD      lui faire

6 PF VT NC,DV EP NC       PF VT NC EP NC
                          j'ai mal au ventre
                          j'ai mal au pied
                          j'ai mal au genou

6 PS,DV VT AC             PS VT AC
                          en a deux
                          en avait deux
                          en a trois

6 PS,DV VT DV,DG EP NC    PS VT DV EP NC
       3 PS,DV,EP VT,VA DV,DG,DM EP AQ,NC en a plus de roses
       3 PS,DV,EP VT,VA DV,DG,DM EP NC    en a plus d'école
```

6 PS,DV,EP VT DV	PS VT DV en a encore en a beaucoup
6 PS,DV,EP VT DV,DG	PS VT DG en a plus
5 CS,DV PI PS VT DN NC 4 CS,DV PI DN NC,VT DN NC 1 CS,DV PI DN VT DN NC	DV PI PS VT DN NC comment on les fabrique les bols? si on l'allume le feu
5 DN,PS NC,VT EP NP	DN NC EP NP la montre pour Maman la montre de Papa le peigne de Philippe
5 DV,DG NC	DV NC plus lait plus pneu plus musique
5 DV,EP NC	EP NC en camion en route en zigzag
5 EP DN,PS NC,VT	EP DN NC avec les plantes avec l'aimant dans le lit par la porte
5 PF PS,EP VT DV,DG	PF PS VT DV j'en veux plus

Chapter V

DEVELOPMENTAL MODELS

One of the most significant and important topics in developmental psychology is that of the language development of the child. There exists a large literature on the subject, and many interesting examples of the acquisition of particular language skills, either of comprehension or of production, have been given. On the other hand, because of what appears to be the bewildering complexity of the language usage taken as a whole, even of a fairly young child, there have been few if any attempts to test systematic models of language development. It should be apparent that the kind of probabilistic grammar that we have constructed for Philippe's speech provides the sort of quantitative framework within which it is possible to conceive and test specific mathematical or formal models of language development. Ultimately, of course, we all desire a completely detailed theory of language acquisition for the child, a theory that is rich enough in its information-processing capacities to provide a detailed account of the child's language development in all major features. We are not prepared to offer anything of this detail or depth in the present context, but we think there are still significant opportunities to formulate and test global developmental models that have some theoretical basis.

Because of the conceptual interest in differentiating between

language development occurring in discrete stages as opposed to an incremental and continuous fashion, we have chosen to test alternative models that represent in a global manner these two ways of thinking about development.

Before entering into any details, it is important to recognize that in either a discrete stage approach or a continuous, incremental approach we must take account of the obvious fact that all normal children develop new language capacities and new skills as they get older, in an especially striking way in the period running from approximately 24 months to 48 months. The intellectually interesting task is not to affirm this obvious fact, but rather to distinguish whether the concept of stages or the equally intuitive concept of continuous development provides a better account of the kind of detailed data we have collected in the case of Philippe.

Secondly, although the data analysis we present is in some respects rather massive, we regard our own efforts as very much preliminary in character. Perhaps the most important reason for saying this is that the data do not show the kind of smoothness properties we would need to test decisively the choice between the two models. The fit to the corpus of more than 15,000 utterances of either class of models we consider is rather bad. The real point of our analysis is to show how one can at least begin to talk about the language development of a child by looking at the entire systematic grammar of his speech, and not merely at examples of individual utterances. On this point we

do not want to be misunderstood. We think that it will continue to be of value to look at individual utterances and to extract from them insights into particular changes in the child's speech. At the same time, it is our thesis that is valuable to try to analyze in a more global fashion the character of broad changes in the grammatical patterns, and other linguistic patterns as well, such as semantic functions.

In considering the overall development of Philippe's grammar during the period under study in this report, it is perhaps natural to begin by asking what sort of curve we get for the introduction of the large number of rules in the grammar. An easy way to look at these data is to graph the cumulative curve, with the abscissa being Philippe's age, and the ordinate being the number of rules used in our sample up to a given age. An analysis of this kind is shown in Figure 1.

The kind of analysis exemplified in Figure 1, however, is quite restricted in character. In the first place, we have to be careful in making inferences of a strong character about the time at which rules are introduced, because our sample based on an hour per week is less than one percent of Philippe's speech per week, and in view of the fact that in the later periods the spacing is even more sparse an even smaller percentage of his total speech is being sampled in a given period. Also, it is reasonable to view the introduction of a particular rule as being only of minor importance. Of greater

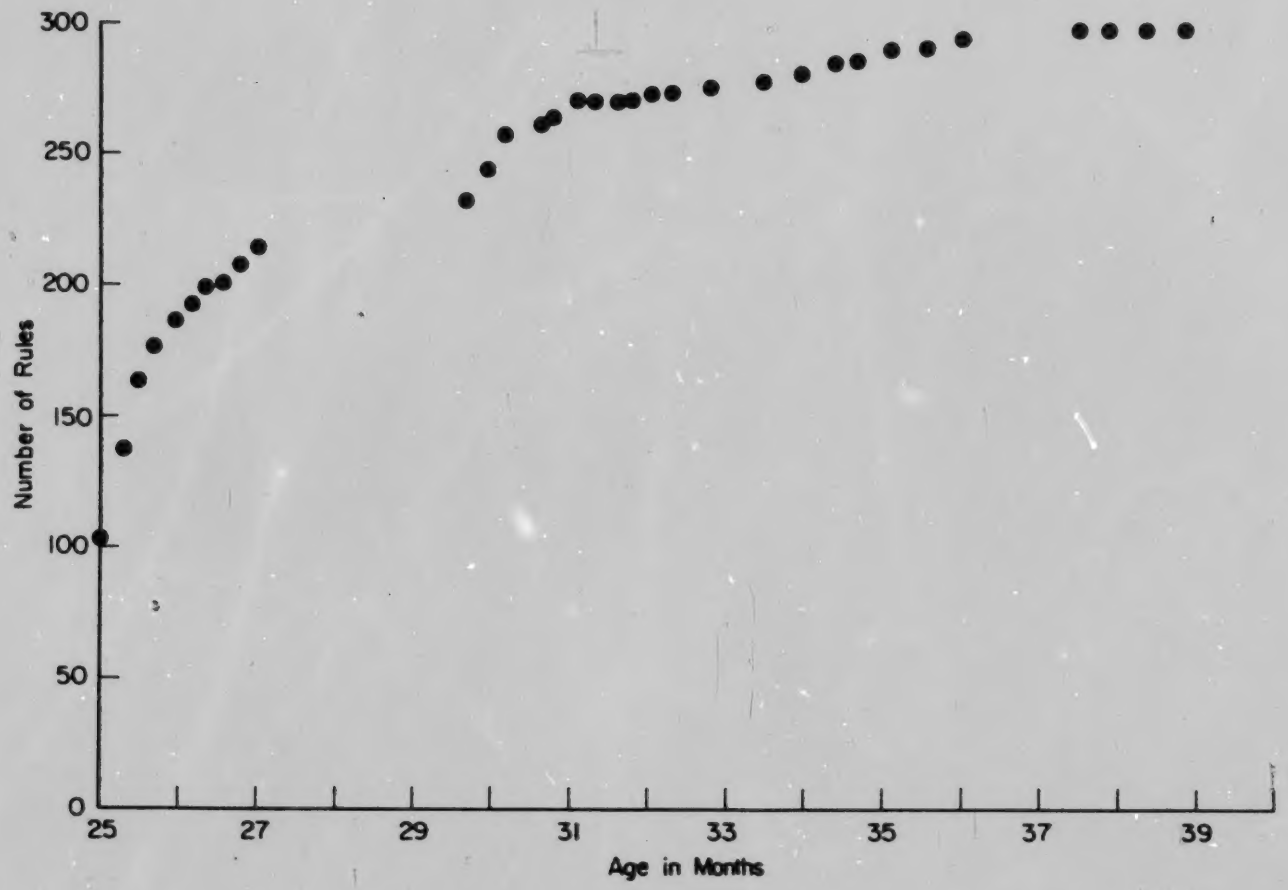


Fig. 1. Cumulative curve of first use of grammatical rules.

importance is the central tendency of his development and the extent to which he continues to use a rule once it has appeared. For the purpose of catching such central tendencies, the probabilistic kind of grammar we have considered earlier seems appropriate and natural.

To pursue this analysis, what we have done is to divide the 33 sessions into six sessions of about an equal number of utterances, with a break being imposed during the long summer vacation in 1971. Even when individual sessions are consolidated into blocks of five and ten sessions, some of the individual rules have a low probability of being used and consequently the behavior of the probabilistic parameters over time can scarcely be studied systematically because of the expected sampling fluctuation being too large relative to the frequency of occurrence of a rule itself.

To meet this problem of small probabilities being assigned to certain rules, as necessarily must be the case when the number of rules in a given group is large, we have merged within each group of rules individual rules into small numbers of classes, these classes themselves being based upon what seem to be relatively intuitive linguistic considerations.

V.1 Merging of the data

Three main principles have guided us in grouping the grammatical rules:

a) Rules which have a low frequency of usage and analyze similar types of utterances should be in the same subgroup;

b) rules of high frequency of usage should be in separate subgroups;

c) rules which are likely to reveal different developmental trends should be in separate subgroups.

GROUP 1 (highest level rules) is divided into 3 subgroups:

1.1 gives the rewrite rules for one and two-words utterances.

1.2 gives the rewrite rules for the noun phrases and adjectival-phrases which are not combined with verbs.

1.3 gives the rewrite rules for verb phrases.

GROUP 2 (incomplete utterances) is divided into 6 subgroups:

1) Rule 2.1 produces adverbs (DV).

2) Rule 2.3 produces adverbs of negation (DG).

3) Rule 2.4 is rewritten as an interjection (KO).

4) Rules 2.5 and 2.6 produce locutions and numerical adjectives, respectively.

5) Rules 2.2, 2.7 to 2.14, 2.16 and 2.17 produce two-word utterances.

6) Rule 2.15 (S1 -> DM DV) generates two consecutive adverbs. This rule has not been used, since DM,DV DV was analyzed as DV DV (which is generated by rule 2.11) after probabilistic disambiguation.

GROUP 3 (noun-phrase and adjective-phrase utterances)
is divided into 5 subgroups:

1) Rule 3.1 derives SN*1 (a non-terminal symbol for noun phrases).

2) Rules 3.5 to 3.9 derive SN*1 (a non-terminal symbol for noun phrases) preceded or not by an adverb, a conjunction of coordination, a locution or an interjection.

3) Rules 3.2 to 3.4 derive SN (a non-terminal symbol for noun phrases) preceded by the expression "un peu de" or followed by an adverb or a vocative.

4) Rules 3.10 to 3.14 all derive SN*1 either preceded by a conjunction of coordination and an adverb, or preceded by a conjunction of coordination and followed by an adverb.

5) Rules 3.15 to 3.20 derive different sorts of adjectival phrases which are not combined with a noun phrase.

GROUP 4 (utterance combination) is divided into 5 subgroups:

1) Rule 4.1 (S3 -> S') derives a non-terminal symbol for utterances with a verb.

2) Rule 4.3 to 4.8, 4.12, 4.14, 4.15 all derive utterances which contain a verb.

3) Rules 4.9, 4.10 and 4.16 derive utterances which contain an adverb of negation.

4) Rules 4.2, 4.11 and 4.13 derive utterances beginning with a conjunctive locution, a conjunction of coordination followed by an adverb, or a conjunction of coordination followed by two adverbs.

5) Rule 4.17 derives direct interrogations (but not all of them, as has been explained earlier in Chapter III.4, rule 4.17).

GROUP 5 (utterances with a verb) is divided into 7 subgroups:

1) Rule 5.1 derives verb phrases without subject.

2) Rule 5.2 derives verb phrases with a subject.

3) Rule 5.3 derives two clauses linked together by a conjunction of subordination.

4) Rule 5.4 derives two independent clauses linked by a conjunction of coordination.

5) Rules 5.5 and 5.6 introduce relative pronouns in the body of a sentence.

6) Rule 5.7 derives sentences in which a prepositional noun phrase is located before the subject and the verb.

7) Rule 5.8 derives utterances in which the verbal group is the object of the noun phrase that precedes it.

GROUP 6 (nominative predicate noun phrases)
is divided into 4 subgroups:

1) Rule 6.1 generates noun phrases that do not begin with a preposition.

2) Rule 6.2 generates noun phrases that begin with a preposition.

3) Rule 6.3 generates personal pronouns that are nominative predicate.

4) Rule 6.4 generates demonstrative pronouns that are nominative predicate.

GROUP 7 (subject noun phrases) is divided into 5 subgroups:

1) Rule 7.1 generates noun phrase subjects.

2) Rule 7.2 generates personal-pronoun subjects.

3) Rule 7.5 generates demonstrative-pronoun subjects.

4) Rule 7.3 (SNS -> SN* PF), Rule 7.4 (SNS -> PF PF), Rule 7.6 (SNS -> PD PD), and Rule 7.7 (SNS -> SN* PD) derive sentences in which the first element is in apposition to the second one, which is the subject of the verb.

5) Rule 7.8 (SNS -> PD DV PD), and Rule 7.9 (SNS -> SN* DV PF), derive utterances in which first element followed by an adverb is in apposition to the subject of the verb.

GROUP 8 (object noun phrases) is divided into 3 subgroups:

1) Rule 8.1 derives noun phrases that do not begin with a preposition and are objects of verbs.

2) Rule 8.2 derives noun phrases that begin with a preposition and are objects of verbs.

3) Rules 8.3 and 8.4 derive interrogative and demonstrative pronouns that are objects of verbs.

GROUP 9 (basic noun phrases) is divided into 4 subgroups:

1) Rule 9.1 is the basic rewrite rule for noun phrases.

2) Rules 9.3 SN* -> NP* EP NP*

9.6 SN* -> NP* EP PT
9.7 SN* -> NP* EP PR

3) Rules 9.2 SN* -> NP* CC NP*
9.4 SN* -> NP* CC NP* EP NP*

4) Rule 9.5 SN* -> NP* EP EP NP*

GROUP 10 (prepositional noun phrases) is divided into 6 subgroups:

1) Rule 10.1 SN** -> EP NP*

2) Rule 10.2 SN* -> EP NP* EP NP*

3) Rule 10.3 SN** -> EP NP* CC NP*

4) Rule 10.4 SN** -> EP NP* CC EP NP*

5) Rule 10.5 SN** -> EP EP NP*

6) Rules 10.6 SN** -> EP PT
10.7 SN** -> EP PR

GROUP 11 (determiner introduction) gives the detailed rewrite rules for noun phrases. It is divided into 9 subgroups:

1) The common characteristic of rules 11.1 to 11.3 is that they have an adjective in pre-position and an adjective in post-position.

2) Rules 11.4 to 11.7 have an adjective in post-position.

3) Rules 11.8, 11.9, 11.11, 11.22, 11.23 and 11.25 have an adjective in pre-position whether preceded or not by a numerical adjective.

4) Rule 11.10 NP* -> det ADJP NC.

5) Rule 11.12 and 11.13 generate noun phrases with numerical adjectives.

6) Rule 11.14 NP* -> DET NC.

7) Rule 11.15 generates nouns and pronouns without determiners and adjectives.

8) Rules 11.21 and 11.24 generate simple noun phrases with only a determiner and a noun.

9) Rules 11.16 to 11.20 generate noun phrases with indefinite articles and determiners used with numerals alone.

GROUP 12 (noun-phrase utterances) presents the rewrite rules of noun phrases that stand by themselves, without a verb; it is divided into 9 subgroups:

1) 12.1 SN*1 -> NP*

2) 12.8 SN*1 -> PF (personal pronoun)
   12.9 SN*1 -> PD (demonstrative pronoun)

3) 12.2 SN*1 -> NP* CC NP*
   12.5 SN*1 -> NP* CC EP NP*

4) 12.3 SN*1 -> NP* EP NP*
   12.6 SN*1 -> NP* EP PR
   12.7 SN*1 -> NP* EP PT
   12.10 SN*1 -> PD EP NP*

5) 12.4 SN*1 -> NP* EP EP NP*

6) 12.11 SN*1 -> EP NP*

7) 12.17 SN*1 -> EP PT
   12.18 SN*1 -> EP PR
   12.19 SN*1 -> EP PD

8) 12.12 SN*1 -> EP NP* EP NP*
   12.15 SN*1 -> EP NP* EP EP NP*
   12.16 SN*1 -> EP EP NP*

9) 12.13 SN*1 -> EP NP* CC NP*
   12.14 SN*1 -> EP NP* CC EP NP*

GROUP 13 (some pronouns and common and proper nouns) is divided into 3 subgroups:

1) Rules 13.1 to 13.3 generate possessive and indefinite pronouns.

2) Rule 13.4 generates common nouns.

3) Rule 13.5 generates proper nouns.

GROUP 14 (adjective phrases) is divided into 2 subgroups:

1) Rule 14.1 generates adjectives in pre-position.

2) Rule 14.2 generates two adjectives in pre-position.

GROUP 15 (post-position adjectives) has only one rule.

GROUP 16 (determiners) is divided into 5 subgroups, one for each rule.

The 5 rules generate demonstrative, interrogative and possessive adjectives, and definite and indefinite articles.

GROUP 17 (numerical expressions) is divided into 2 subgroups:

1) Rule 17.1 generates cardinal adjectives.

2) Rule 17.2 generates ordinal adjectives.

GROUP 18 (verb phrase structures) is divided into 7 subgroups:

1) Rule 18.1 generates verb phrases (GV).

2) Rule 18.2 generates verb phrases (GV') that begin with a preposition or a personal pronoun.

3) Rule 18.3 introduces auxiliaries.

4) Rules 18.4, 18.5, 18.7, 18.10 and 18.12 generate several sorts of verb phrases that include an auxiliary or a modal.

5) Rule 18.6 generates verb phrases with a modal followed by GV'.

6) Rules 18.8 and 18.9 generate verb phrases with a modal followed by GV.

7) Rules 18.11, and 18.13 to 18.15 generate verb phrases with a personal pronoun before the modal or the auxiliary.

GROUP 19 (auxiliaries) is divided into 2 subgroups:

1) Rule 19.1 AUX1 -> VA (terminal symbol for être and avoir)

2) Rule 19.2 AUX1 -> VA VD (terminal symbol for modals)

GROUP 20 (modals) is divided into 3 subgroups:

1) Rule 20.1 AUX2 -> VD
2) Rule 20.2 AUX2 -> VD VD
3) Rule 20.3 AUX2 -> VD ps VD

GROUP 21 (verbal group) is divided into 13 subgroups:

1) Rules 21.1 to 21.3, 2.5 to 21.8, 21.11, 21.15, 21.16, 21.18, and 21.21 to 21.23 generate verb phrases with a copula.

2) Rule 21.4 GV -> COP SA SN.

3) Rules 21.13, 21.17 and 21.20 generate sentences in which the copula is followed by the verbal group GV'.

4) Rules 21.9, 21.10, 21.12, 21.14, 21.19, 21.29, 21.30, 21.31, 21.33, 21.57, 21.58, 21.68, 21.69, 21.74 and 21.75 generate several sorts of verb phrases in the negative form.

5) Rule 21.36 GV -> V DG.

6) Rule 21.24 GV -> V.

7) Rules 21.27, 21.32, 21.43, 21.44, 21.47 to 21.51, 21.53 to 21.55, and 21.59 to 21.62 generate verb phrases in which the verb has no noun phrase object, but is followed by such words as adverbs, locutions, or numerals.

8) Rule 21.25 GV -> V SNP.

9) Rule 21.26 GV -> V SNP DV.

10) Rules 21.28, 21.35, 21.37 to 21.42, 21.45, 21.46 and 21.56 generate several forms of verb phrases with transitive or intransitive verbs.

11) Rule 21.34 GV -> V SNP SNS.

12) Rules 21.52, 21.67, 21.76 to 21.80 generate sentences in which there is a relative or interrogative pronoun that comes after the verb of the main clause.

13) Rules 21.63 to 21.66 and 21.70 to 21.73 generate verb phrases that begin with the word y.

GROUP 22 (verb phrases with prepositions or personal pronouns) is divided into 13 subgroups:

1) Rule 22.1 GV' -> PS V.
2) Rule 22.2 GV' -> PS V SNP.
3) Rule 22.3 GV' -> PS V DG.
4) Rule 22.4 GV' -> PS V DV.
5) Rule 22.9 GV' -> PS V DV SNP.

6) Rules 22.6 to 22.8 and 22.10 to 22.12 generate verb phrases that begin with a personal-pronoun object of the verb which follows it.

7) Rules 22.5, 22.13, 22.14 and 22.23 generate the same sorts of verb phrases, but in the negative form.

8) Rule 22.15 generates the verb s'en aller.

9) Rule 22.16 GV' -> EP V.
10) Rule 22.17 GV' -> EP V SNP.
11) Rule 22.18 GV' -> EP PS V.

12) Rules 22.19 to 22.22 generate different forms of verb phrases beginning with a preposition.

13) Rules 22.24 and 22.25 generate sentences beginning with pour que which have not been used.

GROUP 23 (intransitive and impersonal verbs) is divided into 2 subgroups:

1) Rule 23.1 V' -> VI (intransitive verbs)
2) Rule 23.2 V' -> VM (impersonal verbs)

GROUP 24 (copula) has only one rule:

24.1 COP -> VC (être, rester, devenir)

GROUP 25 (transitive and pronominal verbs) has 3 subgroups:

1) 25.1 V -> VI (transitive verbs)
2) 25.2 V -> VR (pronominal verbs)
3) 25.3 V -> VN (verbs which can be transitive or intransitive)

GROUP 26 (predicate adjectives) has 2 subgroups:

26.1 SA -> SA'
26.2 SA -> DM SA'

GROUP 27 (adjectives without determiners) has 3 subgroups:

1) 27.1 SA' -> AQ
2) 27.2 SA' -> AQ EP NC
3) 27.3 SA' -> AQ EP AQ NC

GROUP 28 (questions) has 11 subgroups:

1) Rule 28.1 Q -> PR.

2) Rules 28.2 to 21.4 generate questions without verbs.

3) Rules 28.5 to 28.7 and 28.19 generate questions in the following form: c'est quoi?, c'est où?.

4) Rule 28.8 Q -> PR LC PR PF SV.
5) Rule 28.9 Q -> PR LC PF SV.

6) Rules 28.10, 28.11, 28.13 and 28.17 generate questions of the following form: qu'est-ce que?

7) Rule 28.12 Q -> PR LC PY SV
Rule 28.14 Q -> PR LC PR SV.

8) Rule 28.15 Q -> PR PD SV.

9) Rules 28.16, 28.18 and 28.24 generate questions of the following form: qui c'est qui?, c'est qui qui?, c'est où que?, c'est quoi qui?, qui y a?.

10) Rule 28.20 Q -> PD VC PR SN.

11) Rules 28.21 to 28.23 generate questions of the following form: c'est quoi? followed by a noun phrase.

GROUP 29 (adverbs of quantity) has two subgroups, one for each rule.

The two rules of this group generate the expression un peu, and un petit peu, respectively.

On the basis of the classification just described, the 317 rules are merged into 106 subgroups.

V.2 Reduction of the data

For each of the six time sections, and each of the subclasses of rules in each group of rules, we have estimated, in the same statistical fashion as before, the probabilistic parameters to be attached to each class. These probabilistic parameters are the basic

data that we want to account for by alternative models. What we are interested in is the way that these parameters change over time.

It is important to note that by restricting ourselves to such parameters we do not in principle restrict ourselves in advance to any systematic developmental aspects of Philippe's grammar, for we could begin if we wanted with the grammar he uses toward the end of our sample and simply assign probability zero to any of the rules not used during an earlier period of his speech. In other words, the grammar can include any constructions desired and then we can determine from the estimation of the kind of parameters just mentioned whether or not that construction actually occurs.

The real question of interest is not, as we have emphasized, the existence or occurrence on one occasion of a particular construction, but rather the central tendency to use a given construction and the way in which the central tendency changes over time. In many respects, what we are doing can be regarded as a detailed extension of the kind of thing already done by psycholinguists who look at the mean length of utterance (MLU). The speech of almost all children exhibits a systematic pattern of continued increase in the MLU over the period of development covered by the Philippe corpus. Our problem is to determine whether similar systematic tendencies can be determined for the use of various grammatical rules.

V.3 The All-or-None Stage Model

The basic assumptions of the all-or-none stage model are two. First, development is discontinuous and may be represented by a relatively small number of stages. Secondly, within each stage, there is a constant probability p_r of rule r being used. The technical assumption is that these probabilities within a given stage for a given group of rules constitute a multinomial distribution, and thus satisfy assumptions of independence and stationarity.

This all-or-none stage model may be viewed as resulting from imposing special restrictions on more general models of learning drawn from either cognitive psychology or from behavioral psychology, especially stimulus sampling theory. Because the intuitive idea of stages is widely accepted and used, it does not seem necessary here to formulate the model in a more general context and derive it by imposing special restrictions on more general models of learning. It should also be emphasized that we shall not test the assumption of a multinomial distribution with fixed parameters for each rule during a given stage by testing, for example, for independence or stationarity.

The only detailed test we shall consider here is the identification of stages and the comparison of the fit of the stage model to the incremental model described below.

It also should be clear that if we do not limit the number of stages, then for each group of rules the data can be fit exactly by a 6-stage model, namely, we just assign a stage for each of the time

sections and fit each probability without error. Such a model, of course, is not interesting and does not give us any insight into the comparison between stage and incremental models. What we have done is impose the requirement that for each group of rules only two stages of development are to be looked for within the period covered by our data. Thus, for example, if a given group of rules is n in number, then we want to fit 2n - 2 parameters. We subtract off two parameters because at each stage for a given group the sum of the probabilities assigned to each rule or subgroup of rules must sum to 1. If we have n rules or n subgroups of rules and six stages we have in general 6n - n degrees of freedom and with 2n - n parameters we have left a net of 4n - (n+2) degrees of freedom that provide a test for the 2-stage model. We shall not really make use from a statistical standpoint of this number of degrees of freedom; that is, we shall not really be interested in assigning a significance level to the goodness of fit of the models, because the data are in too crude a form and the fit of the models not sufficiently good to warrant a detailed goodness-of-fit investigation.

V.4 Incremental model

A qualitative formulation of the discrete stage model is relatively straightforward and has been outlined above. Matters are more complicated in the case of the incremental model. The most desirable approach is to derive a stochastic differential equation from qualitative considerations, and then to solve this differential

equation to obtain the predicted developmental curve for a given group of grammatical rules.

Without claiming that we are yet in a position to give a definitive qualitative theory of the incremental model, we do believe we can offer postulates that are intuitively sensible at a relatively gross level of approximation. As in the case of many attempts to model a highly complex situation, we introduce probabilistic assumptions that we test only in their mean validity, without any claim to being able to extend the theory to examine in detail individual sample paths.

In the five assumptions that follow a central concept is that of a conducive occasion for a given group of rules to be used. Some such notion is needed because the developmental probabilities for use of a rule are conditional probabilities -- conditional on the use of some one rule of the group to which it belongs. It is apparent from the formulation of the five assumptions that this concept of conducive occasion is taken as primitive, and the fifth assumption makes explicit our probabilistic postulate about the occurrence of such occasions. In our judgment it is a central task of a deeper developmental theory that includes the semantics of context to account for the specific character and occurrence of such occasions. It is not within the power of a purely syntactic developmental theory.

Assumption 1. On the occasion of an utterance the probability is one that the child will try a grammatical rule from a group that is conducive to the occasion.

Assumption 2. Immediately after a rule r is used, from his more developed model of comprehension the child will judge the appropriateness of the best choice of a rule from the given group. This appropriateness is represented in the mean by a constant probability π_r.

Assumption 3. For each rule r of a group there is a linear incremental change in the probability of use on a conducive occasion as a function of the constant probability π_r of its appropriateness. Thus on conducive occasions

$$p(t + h,r) = (1 - \theta)p(t,r) + \theta\pi_r .$$

Assumption 4. The probability of using a rule r is changed only on occasions conducive to use of the group of grammatical rules to which r belongs.

Assumption 5. The occurrence of occasions that are conducive to the child's use of any given group of grammatical rules follows a Poisson law, i.e., the intervals between occurrence of these conducive occasions are independently and identically distributed.

From these five assumptions, we can derive a simple mean stochastic differential equation. First, let μ be the parameter of the Poisson process for the occurrence of occasions conducive to the use of a given group of rules. As already indicated, let p(t,r) be the mean probability of using rule r of a given group at time t.

Thus with probability h we have

$$p(t + h,r) = (1 - \theta)p(t,r) + \theta\pi_r$$

with probability o(h) more than one conducive occasion occurs in the interval h, and with probability $1-\mu h-o(h)$ no such occasion arises at all, and thus by assumption (4)

$$p(t + h,r) = p(t,r) .$$

Combining probabilities and dividing by h, we have:

$$\frac{p(t + h,r) - p(t,r)}{h} = -\theta\mu p(t,r) + \theta\mu\pi_r + \frac{o(h)}{h} .$$

Whence as $h \to o$ we obtain the differential equation

$$\frac{dp(t,r)}{dt} = -\theta\mu p(t,r) + \theta\mu\pi_r ,$$

whose solution for the boundary condition $p(t,r) = p_r$ for $t = t_1$ is:

$$p(t,r) = \pi_r - (\pi_r - p_r)e^{-\alpha(t-t_1)} ,$$

where $\alpha = \theta\mu$.

V.5 Test of the two-stage model

Using the six temporal sections made up from the 33 sessions as described earlier, we have tested the two-stage model individually for each of the 29 groups of rules. Secondly, for each class of rules within a given group, we have estimated two parameters, the probability of use in the first stage and the probability of use in the second stage.

To give the two-stage model the optimal chance of fitting the data, we determined on the basis of the data the best breakpoint between the stages for each class of rules. Our methodology for this is straightforward but to be completely explicit we define the relative quantities. Let $n_{r,j}$ be the number of times that rule r is used in time section j. Then for a given class, the first stage probability $\hat{\rho}_{1,b,r}$ is defined as follows where r is the rule and b is the breakpoint, the breakpoint having the values b = 1,2,3,4,5. It is understood that the notation b = 1, for example, means that the break occurs between the first and second sections.

$$\hat{\rho}_{1,b,r} = \frac{\sum_{j=1}^{b} n_{r,j}}{\sum_{j=1}^{b} \sum_{k \in Class(r)} n_{k,j}}$$

In similar fashion we define the probability in the two-stage model for rule r being used in the second scage:

$$\hat{\rho}_{2,b,r} = \frac{\sum_{j=b+1}^{6} n_{r,j}}{\sum_{j=b+1}^{6} \sum_{k \in Class(r)} n_{k,j}} .$$

We now determine the best breakpoint by computing the sum of squares for each possible breakpoint for the two-step model. We first made this computation for each group of rules. The formula for rule r in a given group with break at b is as follows:

$$S^2(r,b) = \sum_{j=1}^{b} \left(n_{r,j} - \rho_{1,b,r} * \sum_{k \epsilon Class(r)} n_{k,j} \right)^2$$

$$+ \sum_{j=b+1}^{6} \left(n_{r,j} - \rho_{2,b,r} * \sum_{k \epsilon Class(r)} n_{k,j} \right)^2 .$$

To obtain the sum of squares for a given group of rules we simply sum over the set of rules r in that group in the expression for $S^2(r,b)$.

Figure 2 shows the distribution of classes across the breaks, that is, the distribution of classes for the best break in terms of our criterion of least squares. The histogram of Figure 2 is based on 27 groups, because two of the groups, groups 15 and 24, have only a single rule and therefore the single probability at both stages is 1.

If we look for the best break across all the groups of rules, then it is not surprising that this occurs at break 2, that is, between the second and third time sections. As the data given in the collection of the corpus in our first report indicate, the longest break by far in the corpus is during the summer of 1971, which is the break between the second and the third time sections. Consequently, this break is a natural one.

All the same, the fit of the two-stage model to the data is rather poor. For example, the sum of squares for the best break, break 2, is $4.4 * 10^8$. To give a sense of the variation in this fit across the various possible breaks, if the break is made between the first and

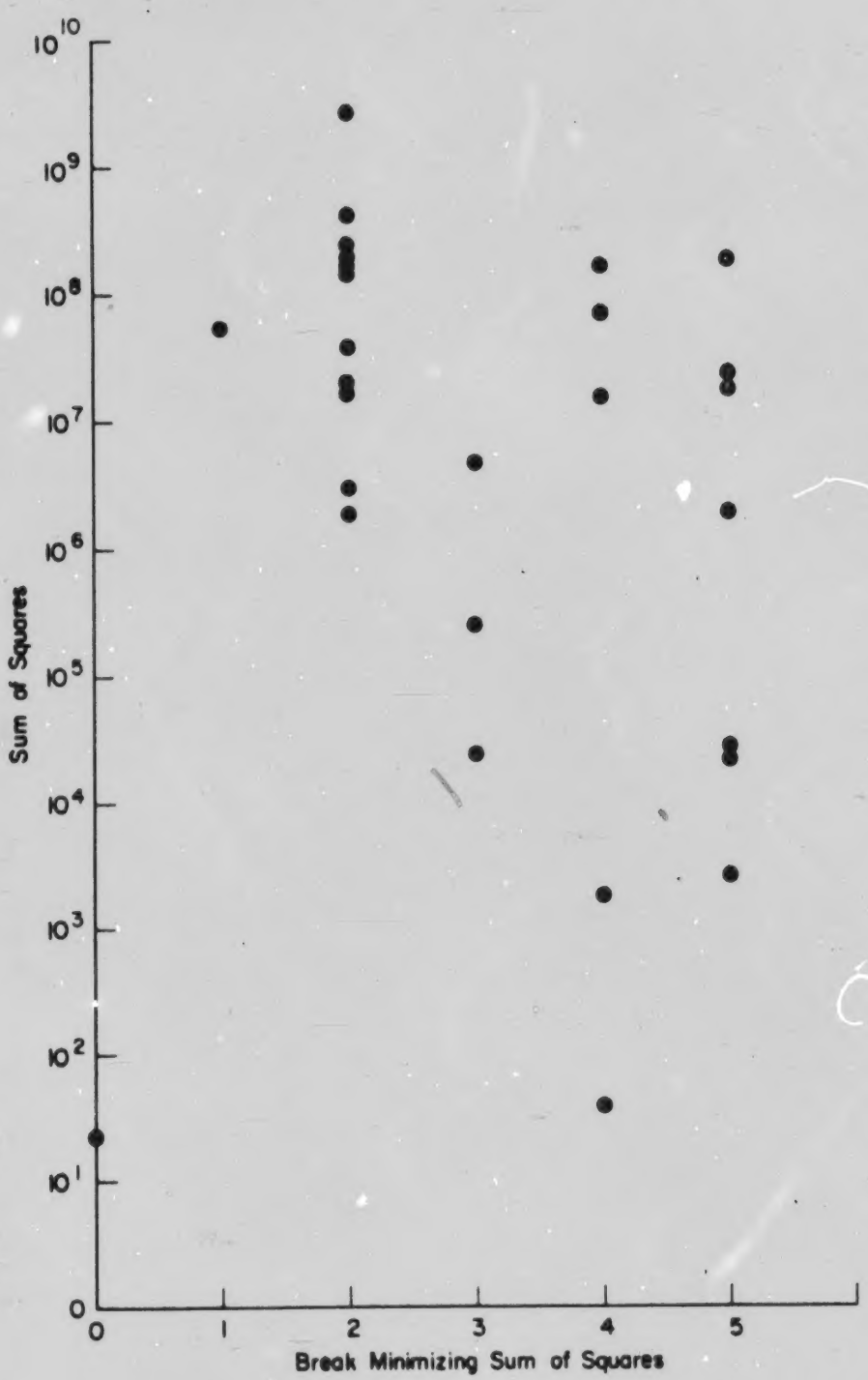


Fig. 2. Distribution of classes for the fit of the two-stage model.

second session, the sum of squares is $8.8 * 10^8$. If it is made in the third break, that is, between the third and fourth time sections, it is $1.2 * 10^9$ and rises from there to $2.0 * 10^9$ for break 5.

We also examined the data using the 33 sections as time sections but there is sufficient instability in the numerical data for the use of rules in a single section not to make these data reliable enough to report. It is worth indicating, however, that looking at the 33 sessions as the time sections the best break occurred between sessions 9 and 10, which is close to the end of time section 2 in the analysis just reported above.

V.6 Fit of the incremental model

The equation for the incremental model requires that for a given group of rules we estimate the parameter α and that for each rule r of a group we estimate its initial probability p_r and its asymptotic probability π_r of appropriateness.

We estimated p_1 by using the probability of use of a rule in the first time section. Because the mean of the first time section is 14 days, we set p(r,14) equal to the observed probability for the first time section.

In the case of π_r and α we used a more complicated procedure. As an initial estimate of π_r we took the probability for the sixth time section, and the preliminary estimate of α was based on the following considerations.

Dropping the subscript r, we may write the basic equation of the incremental model as follows:

$$p(t) = \pi - (\pi - p_1)e^{-\alpha(t-t_1)} ,$$

and thus:

$$\frac{\pi - p(t)}{\pi - p_1} = e^{-\alpha(t-t_1)} ,$$

whence

$$\gamma(t) = \ln \left| \frac{\pi - p(t)}{\pi - p_1} \right| = -\alpha(t - t_1) ,$$

and so

$$\alpha = \frac{\gamma(t)}{t_1 - t} .$$

Our preliminary estimate of α_r was then

$$\hat{\alpha}_r = \frac{1}{4} \sum_{j=2}^{5} \frac{\gamma(j)}{t_1 - t_j} ,$$

using time-sections 2, 3, 4 and 5. We then obtained the initial estimate $\hat{\alpha}$ for a group by simply averaging over the $\hat{\alpha}_r$'s.

Starting with these initial estimates we then minimized the sum of squares $S^2(\alpha, \pi_r, r)$ for each rule of a group by searching over the intervals defined by

$$\hat{\pi}_r \pm .1\hat{\pi}_r$$

$$\hat{\alpha} \pm .1\hat{\alpha} .$$

Final estimates of an accuracy appropriate to the observed data were obtained by this method. The various figures considered below that compare the observed data and theoretical curves directly support this conclusion.

Before turning to the detailed examination of some of the developmental curves for individual rules, it is worth noting that the sum of squares for the incremental model is $1.4 * 10^5$, which is considerably less than the sum of squares for the two-stage model and is indicative of a definitely better fit.

We now turn to some detailed observations about the fit of the incremental model to various groups of rules.

The three very general production rules of group 1 are used at a high level in any syntactic tree that generates an utterance. Rule 1 is the production rule for one- and two-word utterances and this production rule shows an increase over the developmental period, whereas the second rule, which is the rewrite rule for utterances that consist of noun phrases and adjectival phrases that are not combined with verbs would as expected show a decrease in usage. The third rule, which is the rewrite rule for whole sentences as well as verb phrases, has a fairly constant use over the developmental period. These basic facts in the observable data are nicely expressed in the theoretical curves, as may be seen from Figure 3.

The production rules of group 2 generate terminals directly, and especially terminals that lead to partial or fragmentary

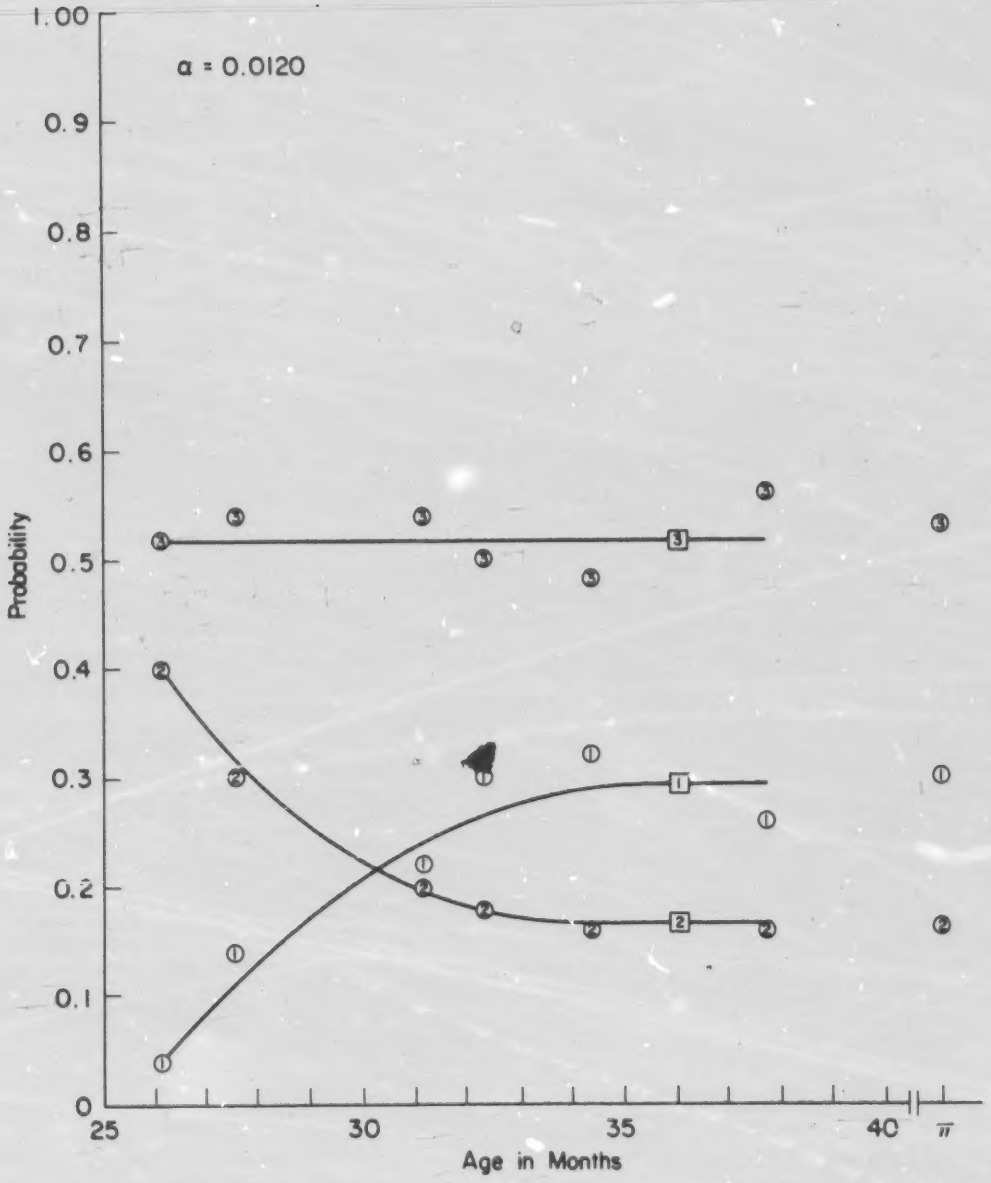


Fig. 3. Fit of incremental model for rules of Group 1, highest level rules.

(The figures displaying the fit of the incremental model all have the following format: the abscissa shows Philippe's age (in months), and the ordinate shows the probability. For each rule-subgroup of the rule class, the encircled numbers show the observed data point. The curves, labeled with corresponding numbers in square boxes, show the theoretical curves predicted by the incremental model. The α value is displayed, and, on the far right, the asymptotic values of the π parameters are given.)

utterances, for example, adverbs standing alone, locutions, interjections, or a preposition followed by an adverb. There are some strong developmental trends evident in the use of the production rules of this group, and they are displayed in Figure 4. For example, as can be seen in Figure 4, there is a very sharp increase in the use of the first production rule, which produces utterances consisting of a single adverb.

The production rules of group 3 generate noun phrases and adjective phrases that are not combined with a verb. The occurrence of noun phrases in the speech of most young children of the age of Philippe is highly frequent, and this is reflected in the observed data as well as in the theoretical curve. In fact, the fluctuations are so slight in the observed data that the theoretical curve is essentially flat, and this is true for all the subgroups of group 3, as shown in Figure 5. As we noted above, the tendency to use a noun phrase as the complete utterance decreases throughout the developmental period.

The production rules of group 4 produce utterances that contain a verb, whether or not the verb is combined with other words, for example, a subject. The first production rule of this group, 4.1, which is the most frequently used, leads to the simplest utterances; it shows a decrease in usage over the period covered by this report. This result is also reflected in the theoretical curve shown in Figure 6. The rules 4.3 through 4.8, 4.12, 4.14, 4.15, grouped for this analysis,

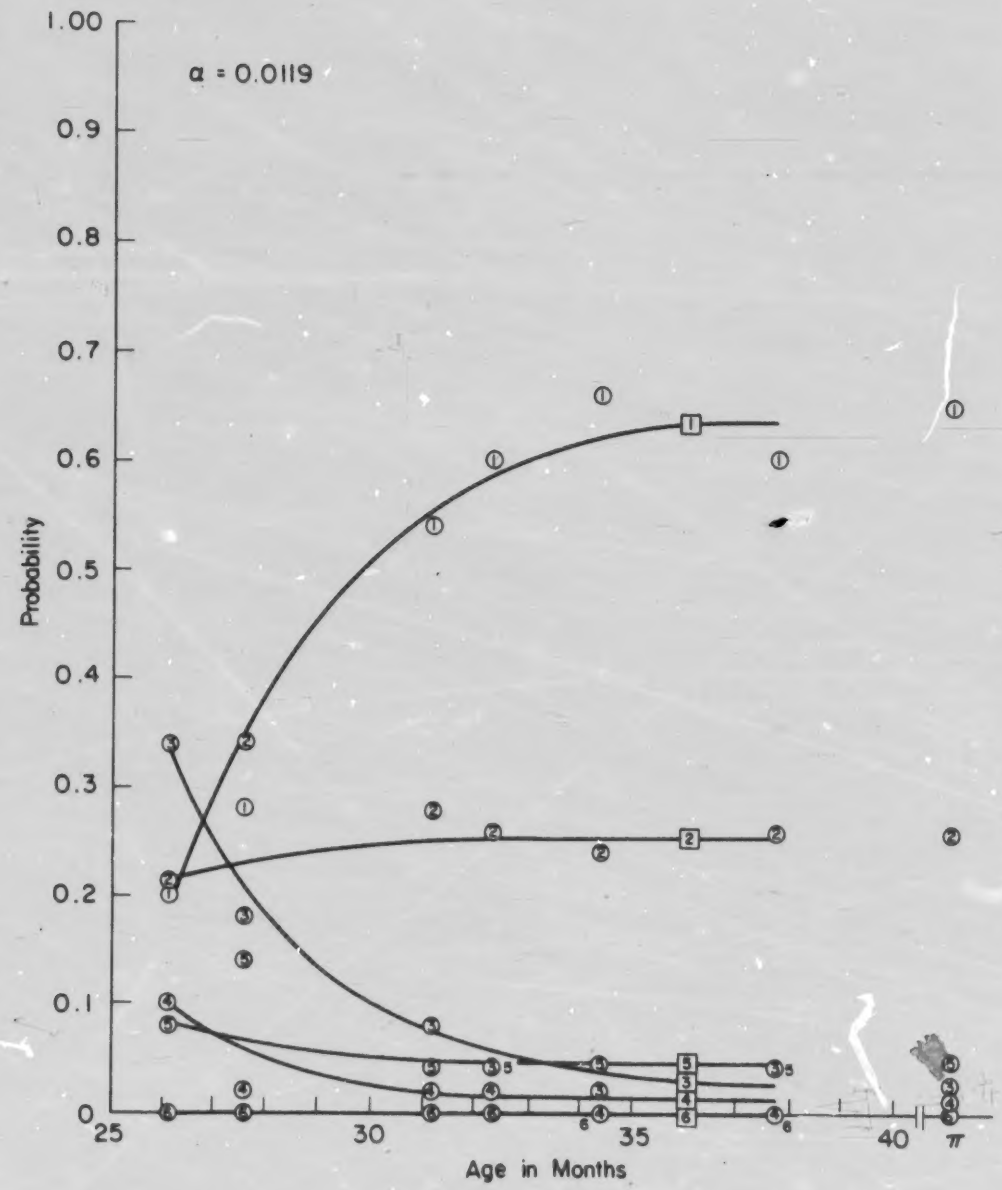


Fig. 4. Fit of incremental model for rules of Group 2, incomplete utterances.

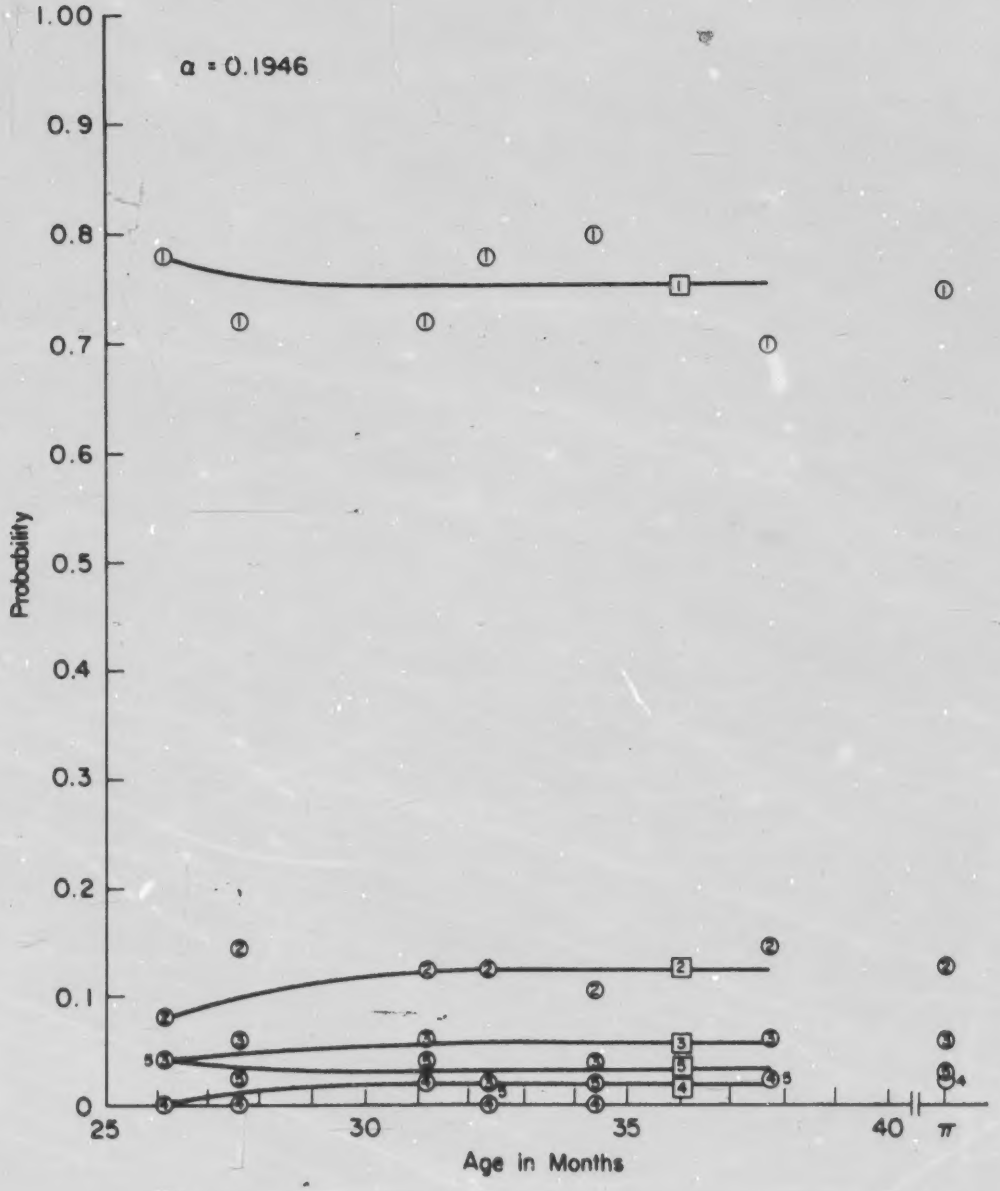


Fig. 5. Fit of incremental model for rules of Group 3, noun phrase and adjective phrase utterances.

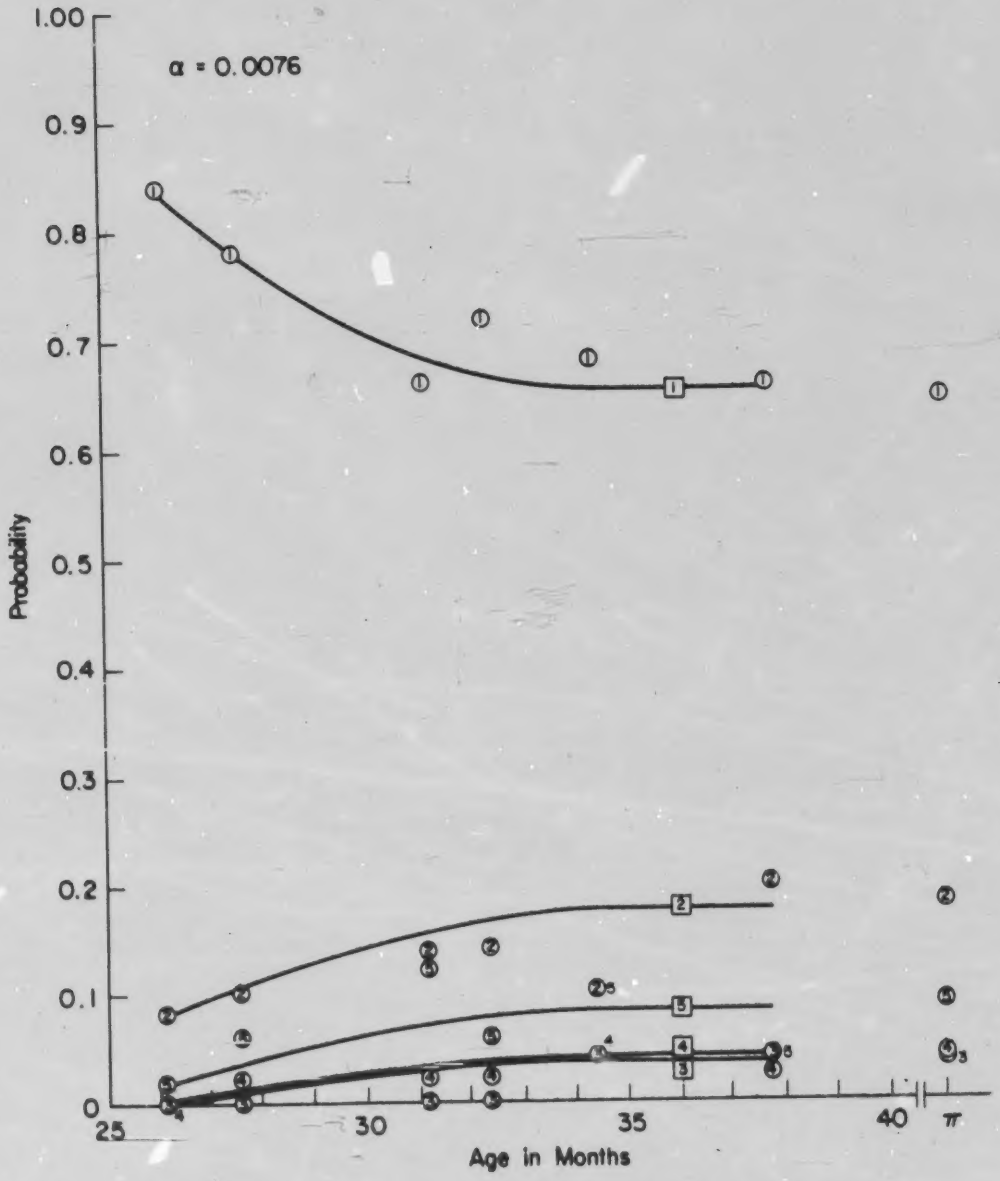


Fig. 6. Fit of incremental model for rules of Group 4, utterance combination.

together represent the tendency to add a word or phrase (such as as locution KO) to the sentence; there is a slight increase in the usage of this subgroup over the developmental period.

The production rules of group 5 begin with the terminal S' to produce utterances that contain a verb. They are extensively used in the corpus. The first rule, which produces verb utterances standing alone, has a sharply decreasing usage in the corpus, ranging from probability of .62 in the first section to .14 in the sixth time section. These data are clearly reflected in the theoretical curve as shown in Figure 7. In contrast, the second production rule of this group is the basic rule for utterances that have a subject as well as a verb, and there is a sharp increase in the use of this production rule from .37 in the first time section to .76 in the last time section, and again the observed changes are reflected in the theoretical curve. In many ways the data and theoretical curves shown in Figure 7 are among the most interesting from a developmental standpoint of the entire collection we are examining. There is a very sharp change in the use of the main production rules of the group during the developmental period being studied. The change has a natural linguistic basis toward increasing complexity of usage and this change, which is incremental and apparently continuous in nature, is fit quite well by the theoretical curves of the incremental model. The other subgroups of group 5 have sufficiently small usage that little comment is required on the very small change throughout the developmental period.

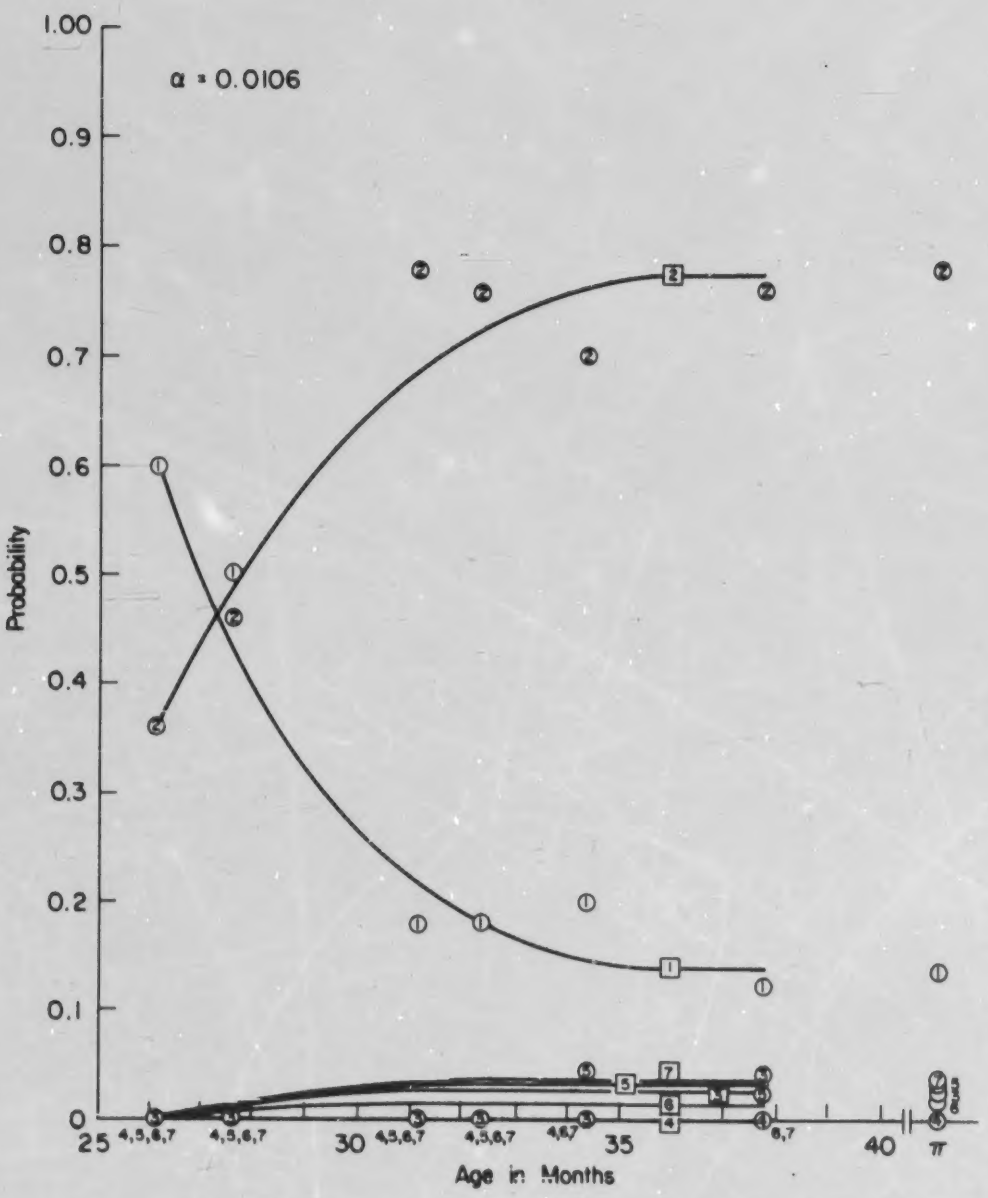


Fig. 7. Fit of incremental model for rules of Group 5, utterances with a verb.

Starting with group 6 the production rules deal with particular subtrees of utterances; for example, groups 6, 7 and 8 are concerned with different non-terminal symbols representing the three main functions of the noun phrases as subject, object and nominative predicate. We present the data for groups 6, 7 and 8 in Figures 8, 9 and 10. It is to be noted that the relative change for group 6 is fairly slight over the developmental period and, of the three groups, is greatest for group 7. Group 7 produces noun phrases as the subject of a verb and what we get is increasing complexity of these subjects over the developmental period as reflected in the relative usage of the production rules. For example, the simple first rule has a sharply decreasing usage and the second rule, which introduces personal pronouns as subject of the verb, has a sharply increasing use. Philippe, as is also the case for English-speaking children, does not have a frequent use of personal-pronoun subjects of utterances in his early speech but uses them increasingly over the period covered by our data. Interestingly enough, the developmental situation is quite different for demonstrative pronouns as subjects of verbs. Here the usage is lower but relatively constant over our developmental period. This may be seen for both the data and theoretical curve of rule 5 of group 7 in Figure 10.

Group 9 consists of rules for rewriting noun phrases. The first rule is the basic rule of this group and dominates the data with little change in relative usage throughout the developmental period, as may be seen from Figure 11.

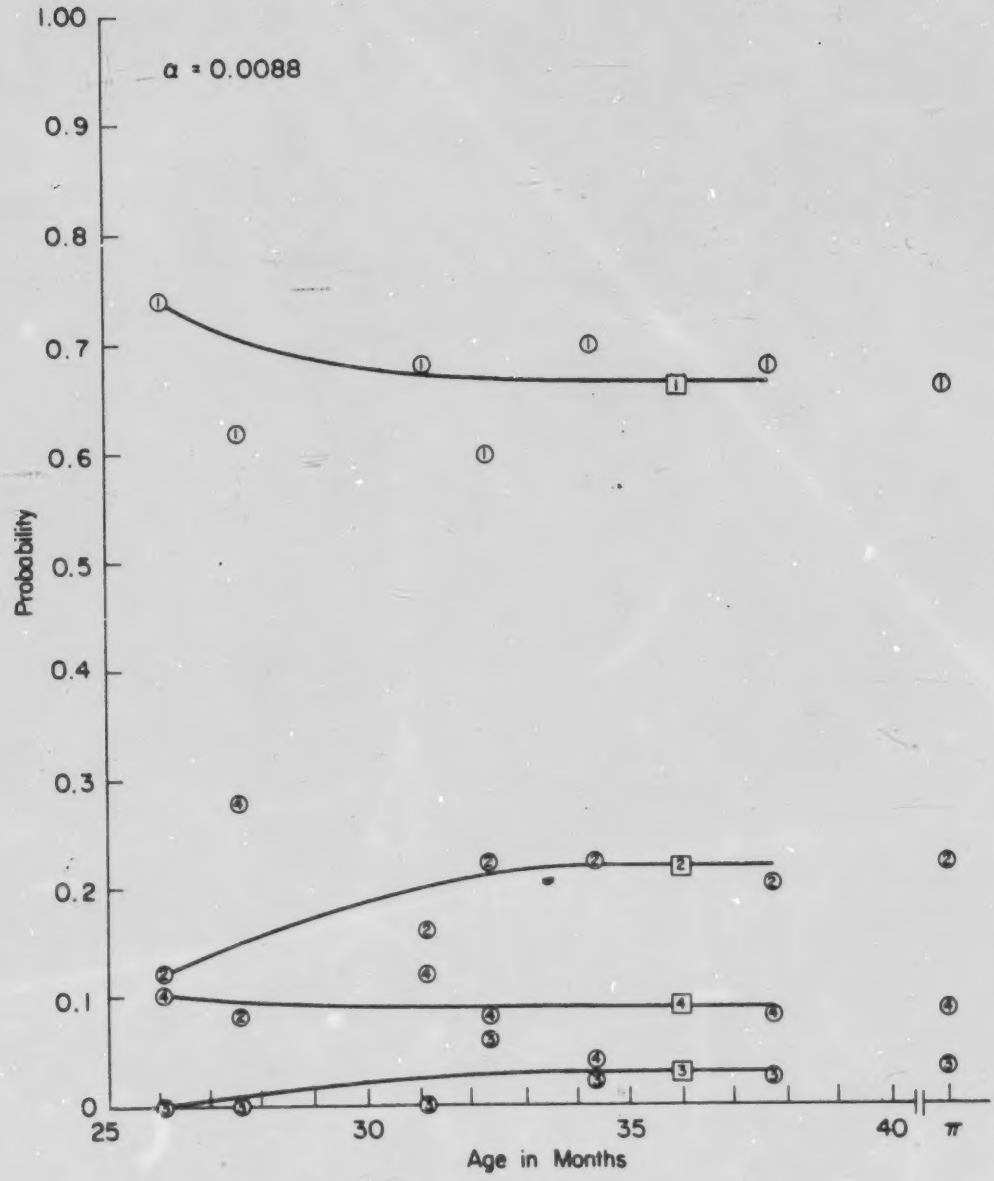


Fig. 8. Fit of incremental model for rules of Group 6, nominative predicate noun phrases.

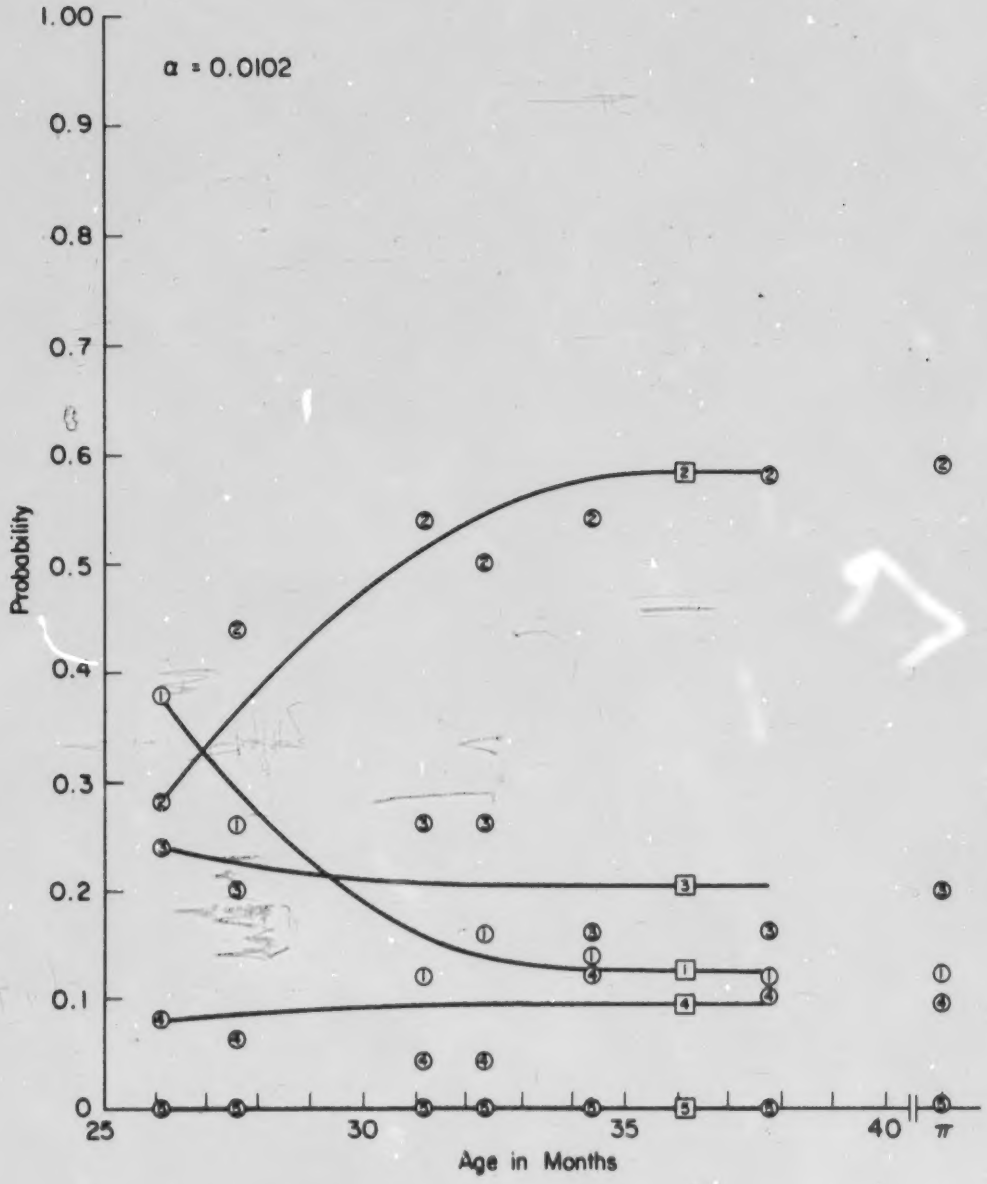


Fig. 9. Fit of incremental model for rules of Group 7, subject noun phrases.

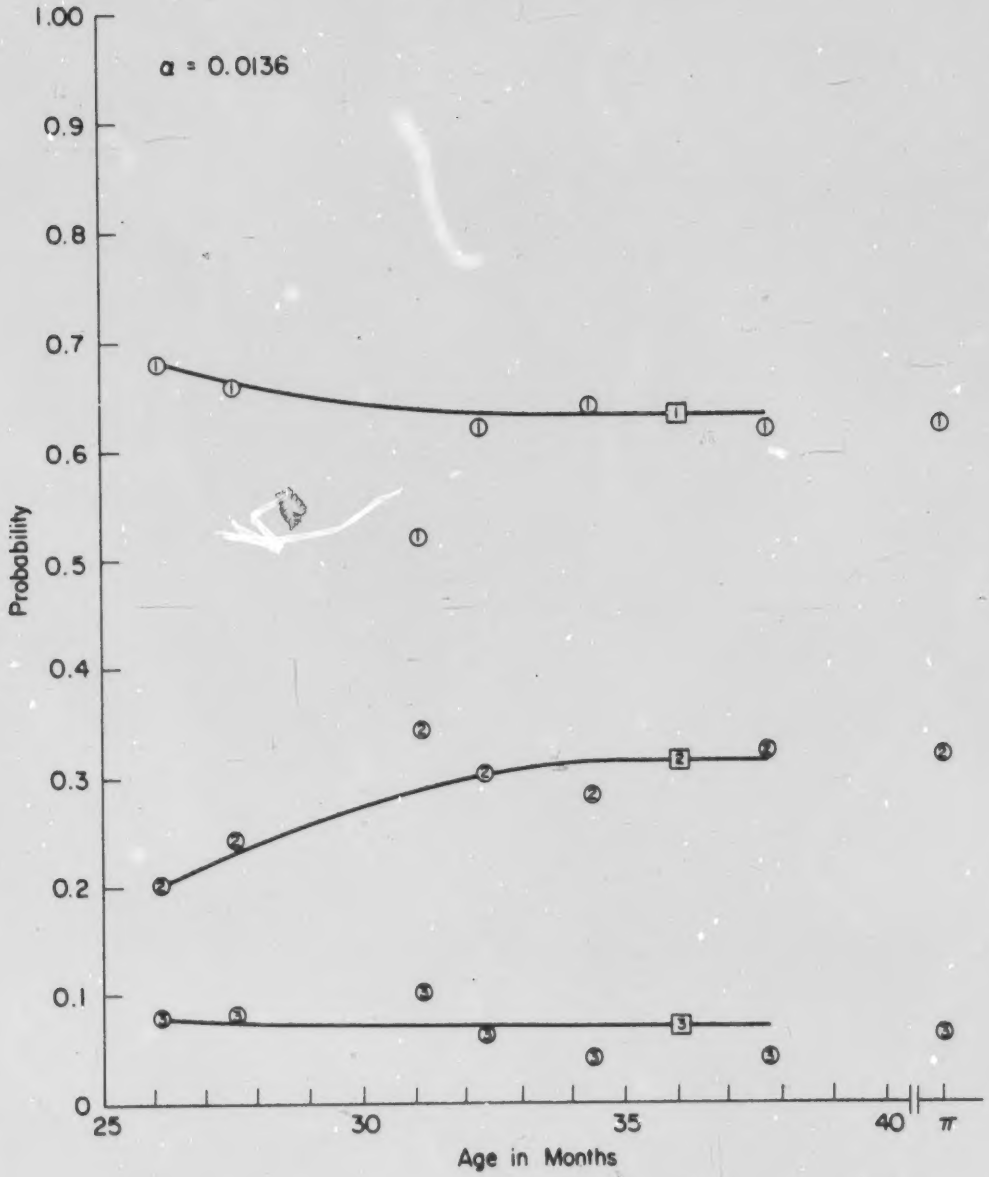


Fig. 10. Fit of incremental model for rules of Group 8, object noun phrases.

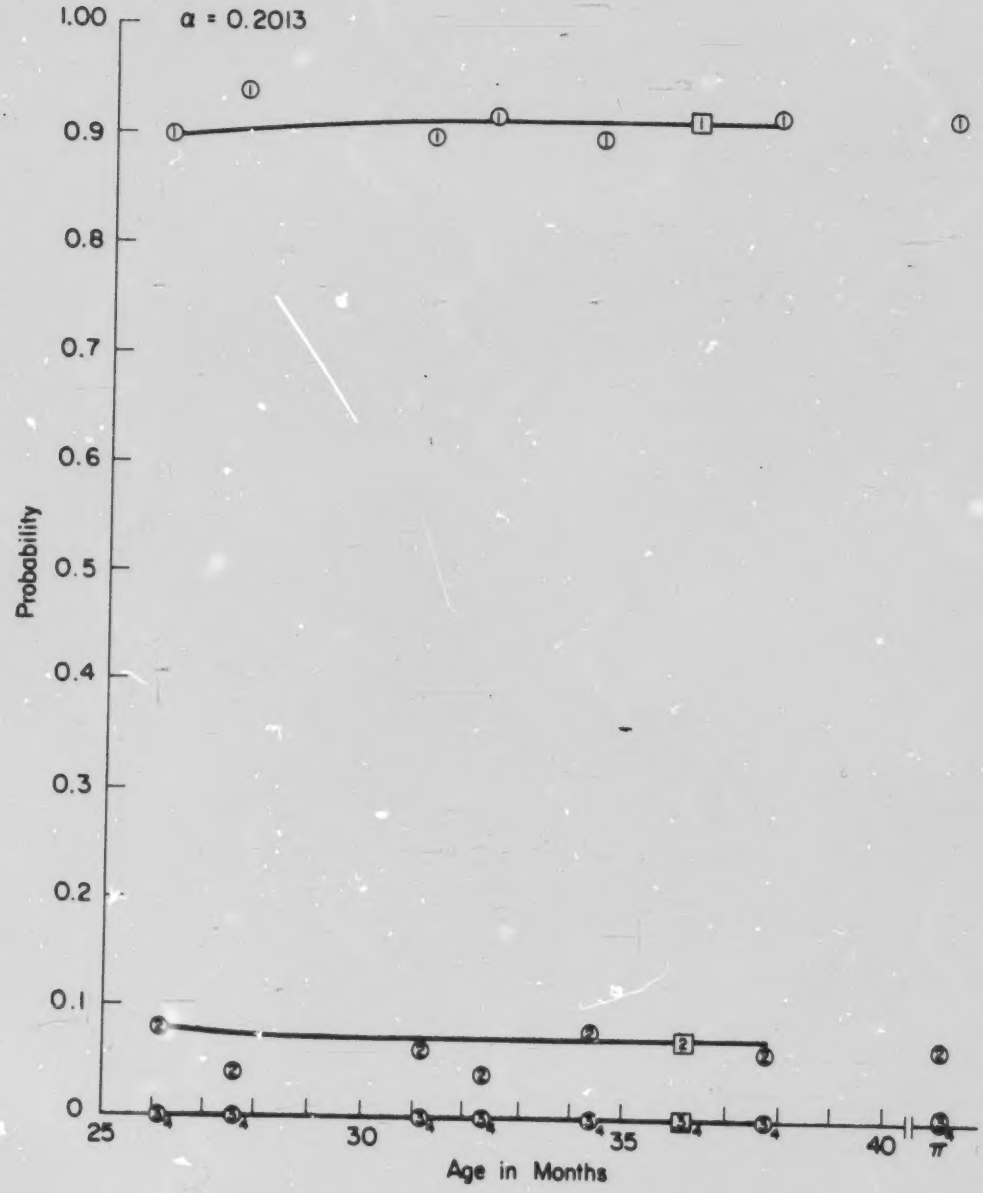


Fig. 11. Fit of incremental model for rules of Group 9, basic noun phrases.

The rules in group 10 introduce noun phrases that begin with EP (plural indefinite articles and prepositions). Figure 12 shows the data and curves for this group; as can be seen, there is little clear development in the data.

The rules in Group 11 introduce determiners, numerical adjectives, and adjective phrases into noun phrases. Rule 11.10, which introduces the determiner and adjective phrase, increases over the period, as does rule 11.14, which introduces the determiner. Rules in the subgroup composed of 11.16 through 11.20, which introduce indefinite articles (AI), also increase slightly. These increases are offset by a clear decrease in rule 11.15, which simply introduces nouns without these modifiers. Other rules and subgroups of rules remain more constant. See Figure 13.

Group 12 (Figure 14) contains rules that combine noun phrases with conjunctions and prepositions. Rule 12.1 introduces the noun phrase without a conjunction or preposition, and it decreases over the period. Rule 12.11 shows a slight increase, although the data are somewhat ambiguous. Rules 12.8 and 12.9 are grouped together, and they show a slight increase.

The rules in Group 13 (Figure 15) introduce the common or proper noun without an article and PI (indefinite pronouns) or PO (possessive pronouns). Rule 13.4 introduces proper nouns, and the use sharply decreases, while rule 13.5, which introduces common nouns, increases. The use of pronouns, with rules 13.1 to 13.3 grouped together, increases over the period.

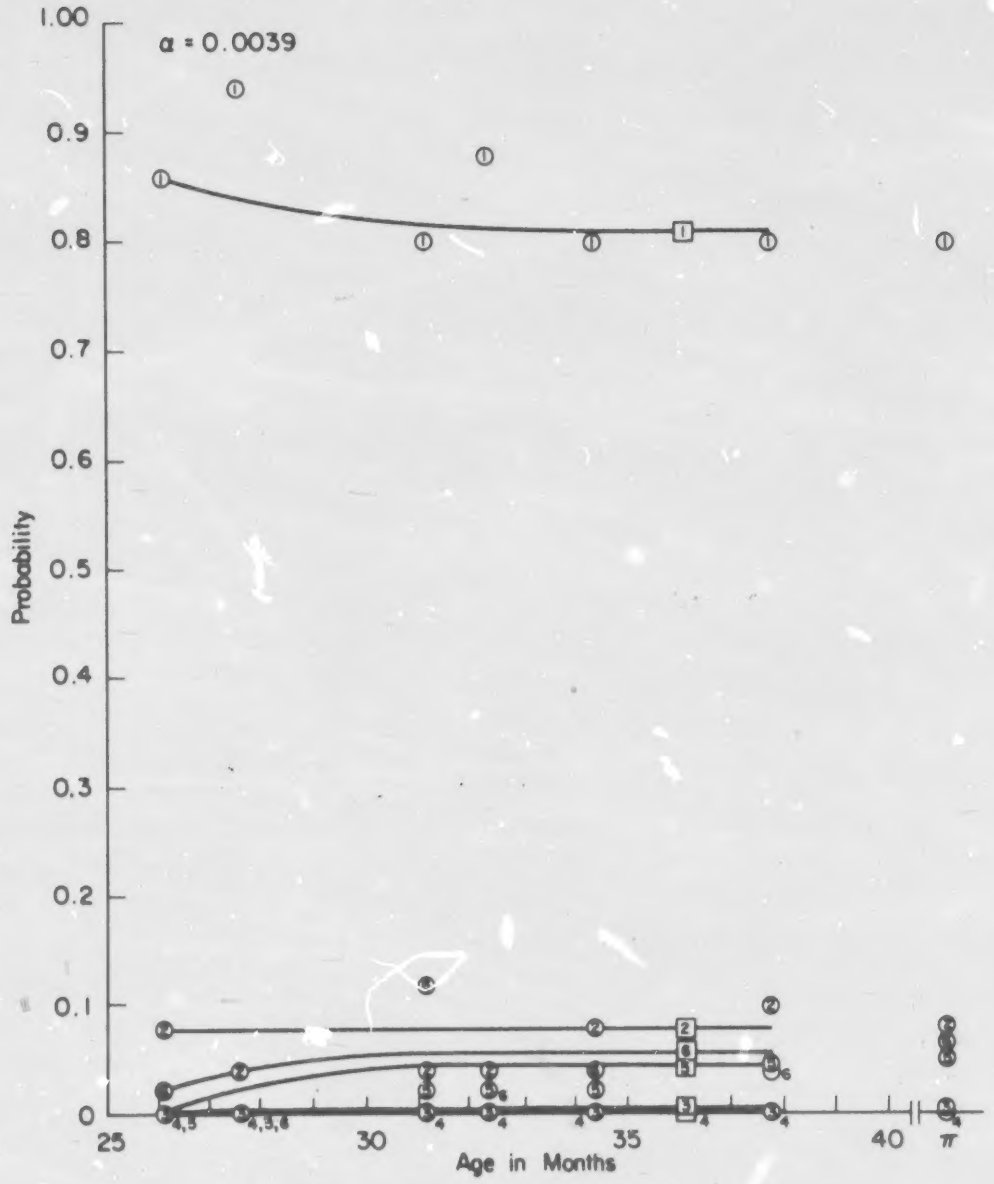


Fig. 12. Fit of incremental model for rules of Group 10, prepositional noun phrases.

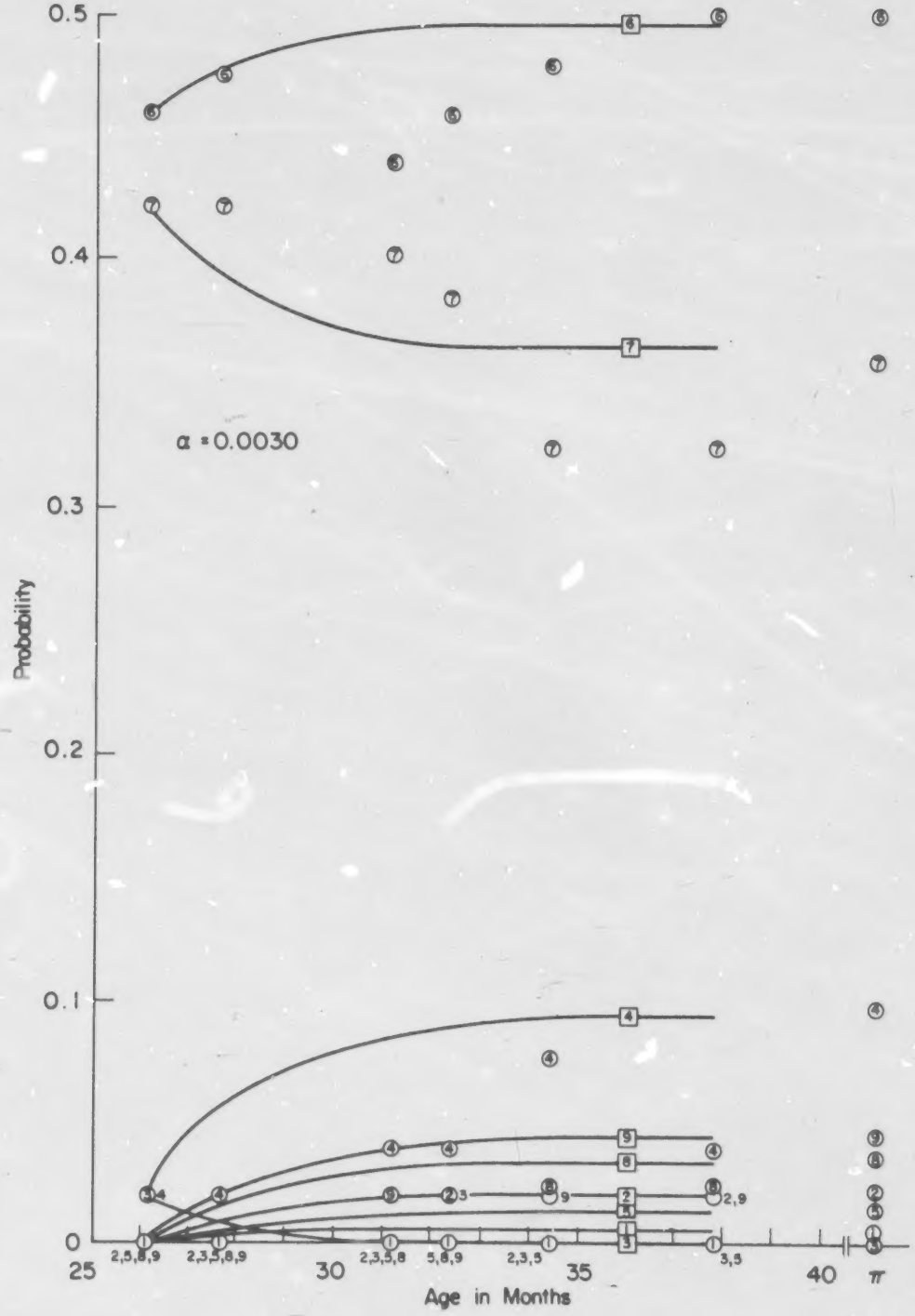


Fig. 13. Fit of incremental model for rules of Group II, determiner introduction.

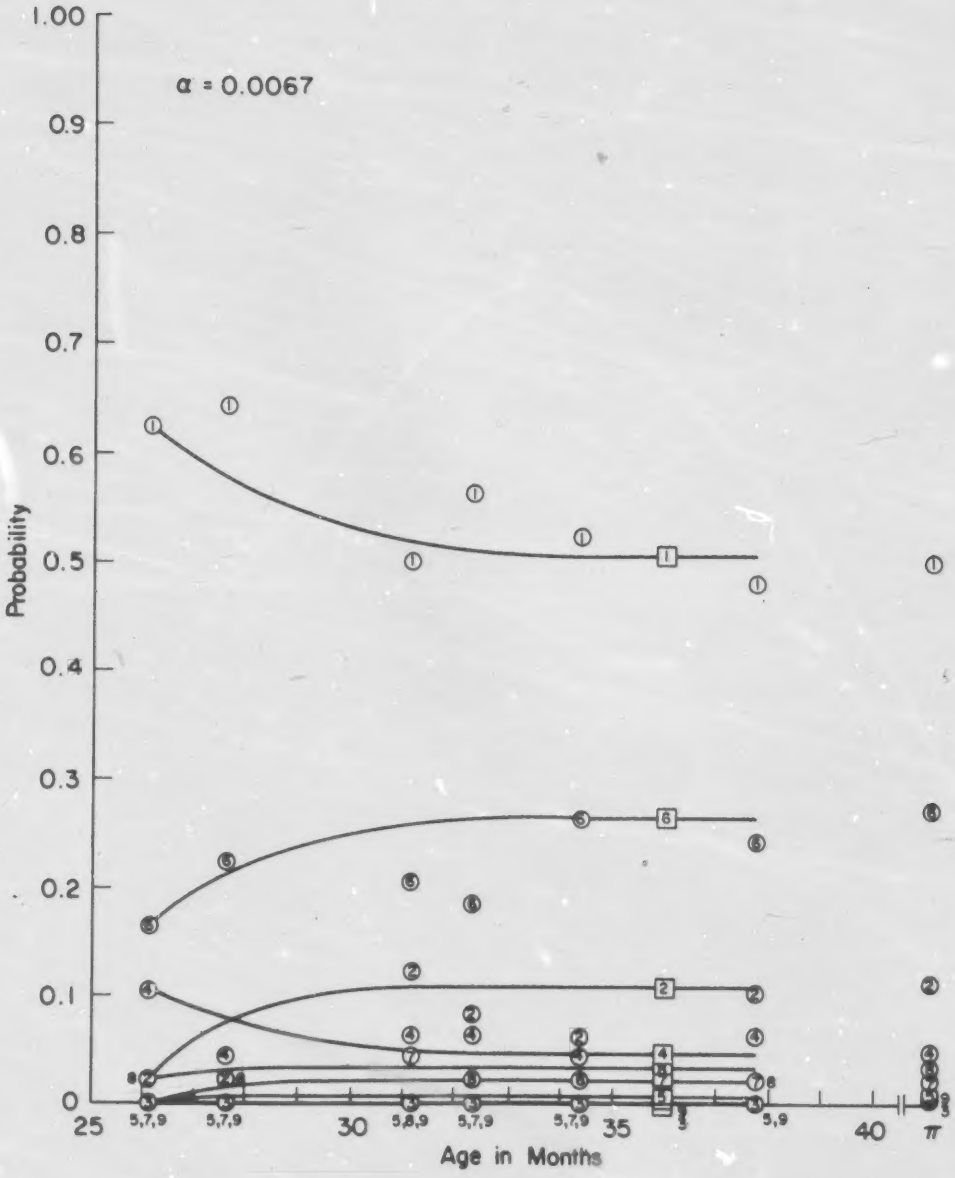


Fig. 14. Fit of incremental model for rules of Group 12, noun-phrase utterances.

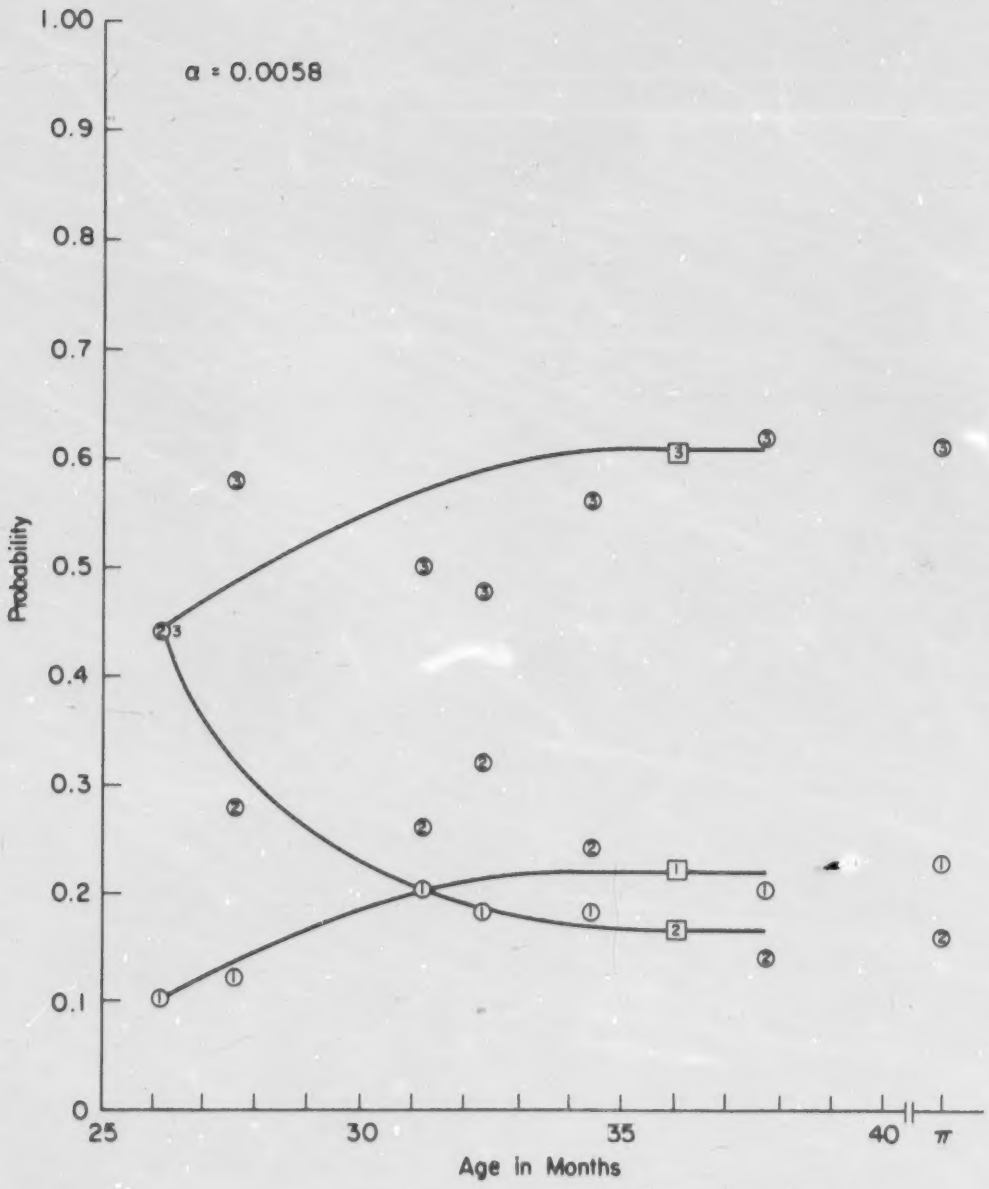


Fig. 15. Fit of incremental model for rules of Group 13, some pronouns and common and proper nouns.

Group 14 (Figure 16) contains two adjective phrase rules, the first of which (14.1) is highly dominant. It shows a slight decrease in use over the developmental period

The rules in Group 16 introduce the determiners. There is a mild increase in the use of demonstrative adjectives (rule 16.1), possessive adjectives (rule 16.3), and indefinite articles (rule 16.5). The definite article shows a marked decrease over the period, as shown in the observed and theoretical data for rule 16.4, in Figure 17.

The rules in Group 17 introduce numerical adjectives. As shown in Figure 18, there is a slight decrease in cardinal adjectives, and an increase in ordinal adjectives. The data and theoretical curves are shown in Figure 18.

Group 18 contains rules that introduce verb phrases. There is a marked decrease in rule 18.1, which introduces the verb phrase GV, and a modest increase in other rules in the group that introduce GV with modifications and GV'. The theoretical curves match the observed data rather well, as shown in Figure 19.

Groups 19 and 20 introduce auxiliary and modal words into verb phrases. The theoretical and observed data are shown in Figures 20 and 21. There is little change over the developmental period.

Groups 21 and 22 introduce the verb phrases GV and GV' respectively. They are large groups of rules, which we divided into 13 subgroups each. Unfortunately, the theoretical curves and observed data do not fit very well on the whole, as seen in Figures 22 and 23.

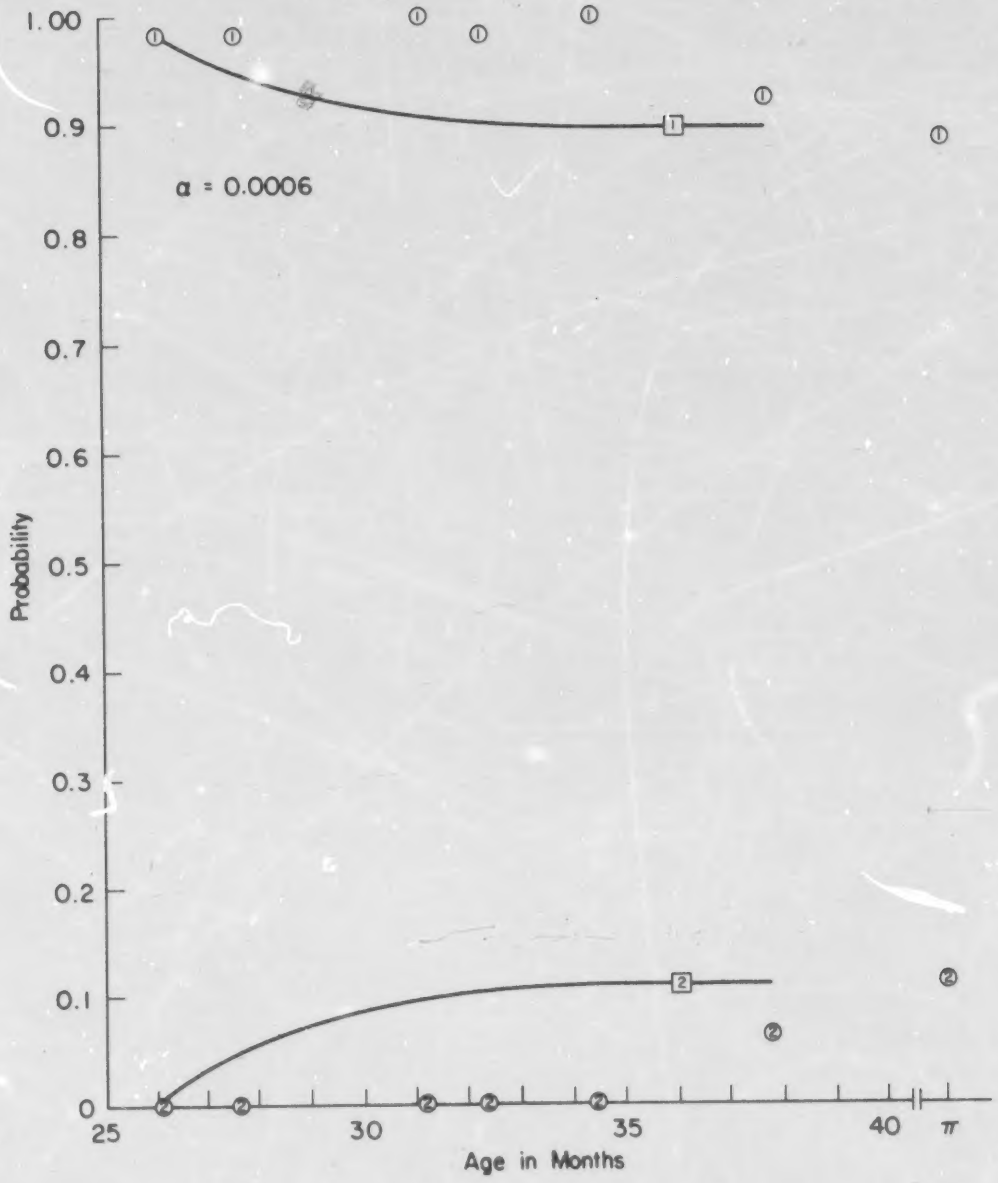


Fig. 16. Fit of incremental model for rules of Group 14, adjective phrases.

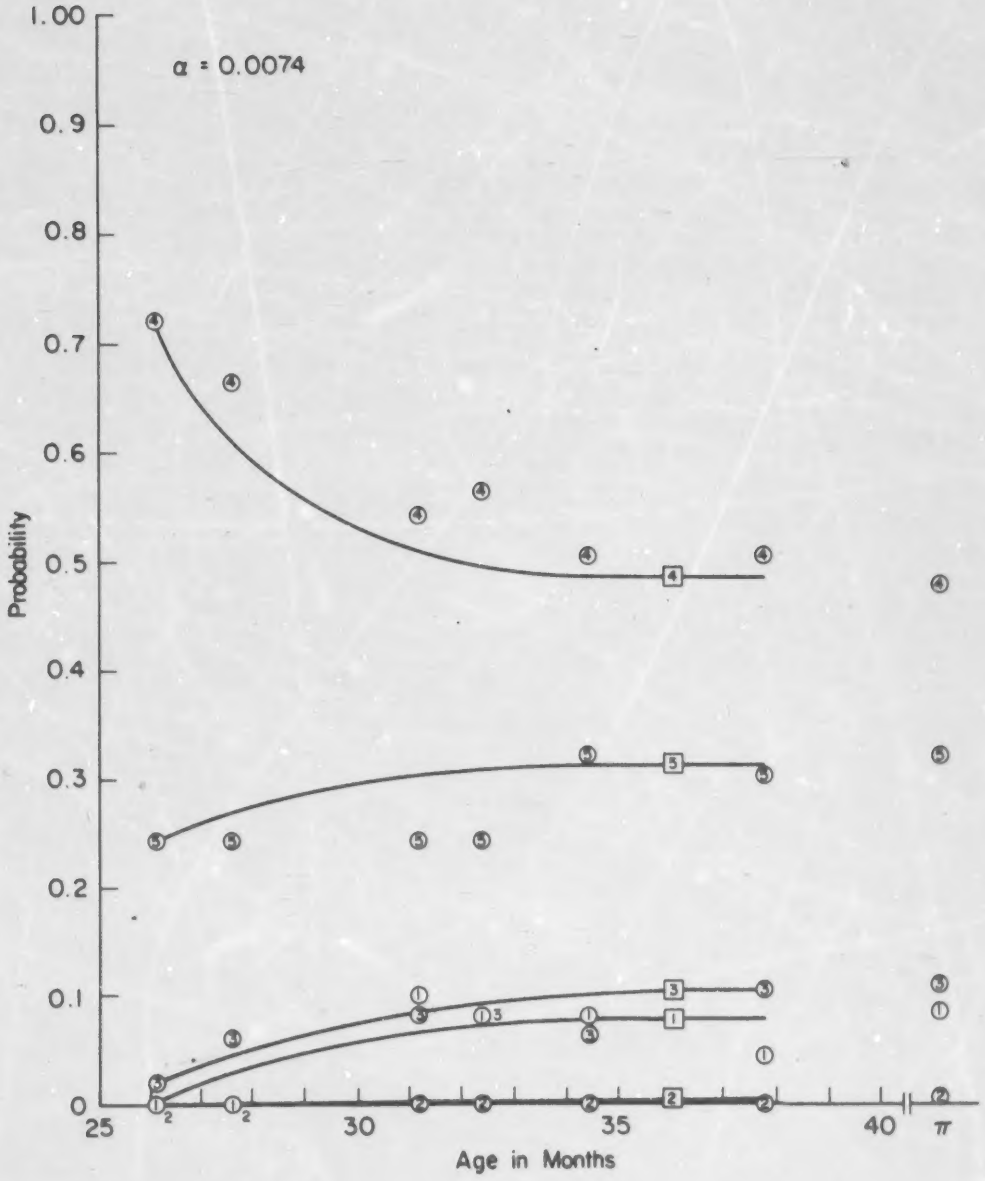


Fig. 17. Fit of incremental model for rules of Group 16, determiners.

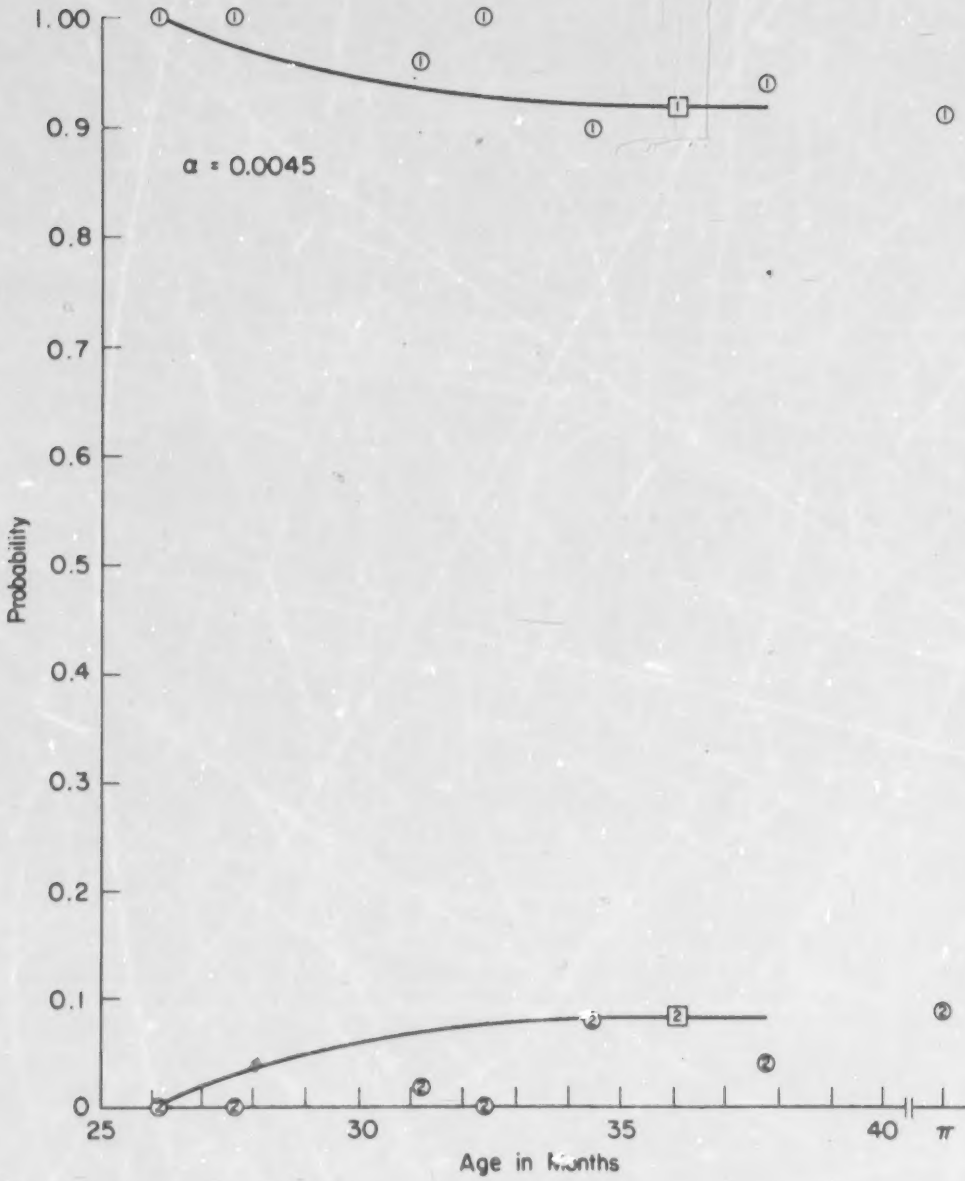


Fig. 18. Fit of incremental model for rules of Group 17, numerical expressions.

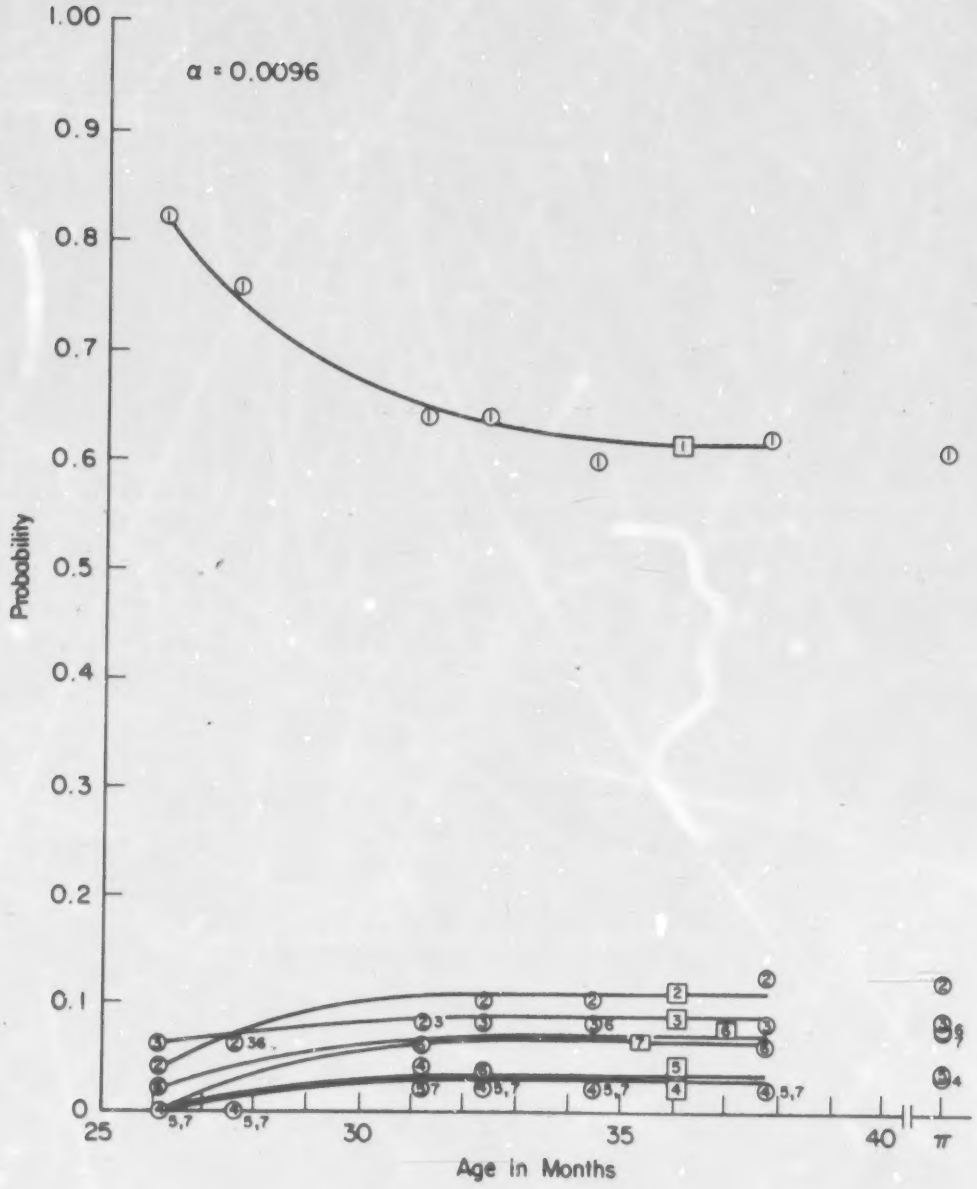


Fig. 19. Fit of incremental model for rules of Group 18, verb phrase structures.

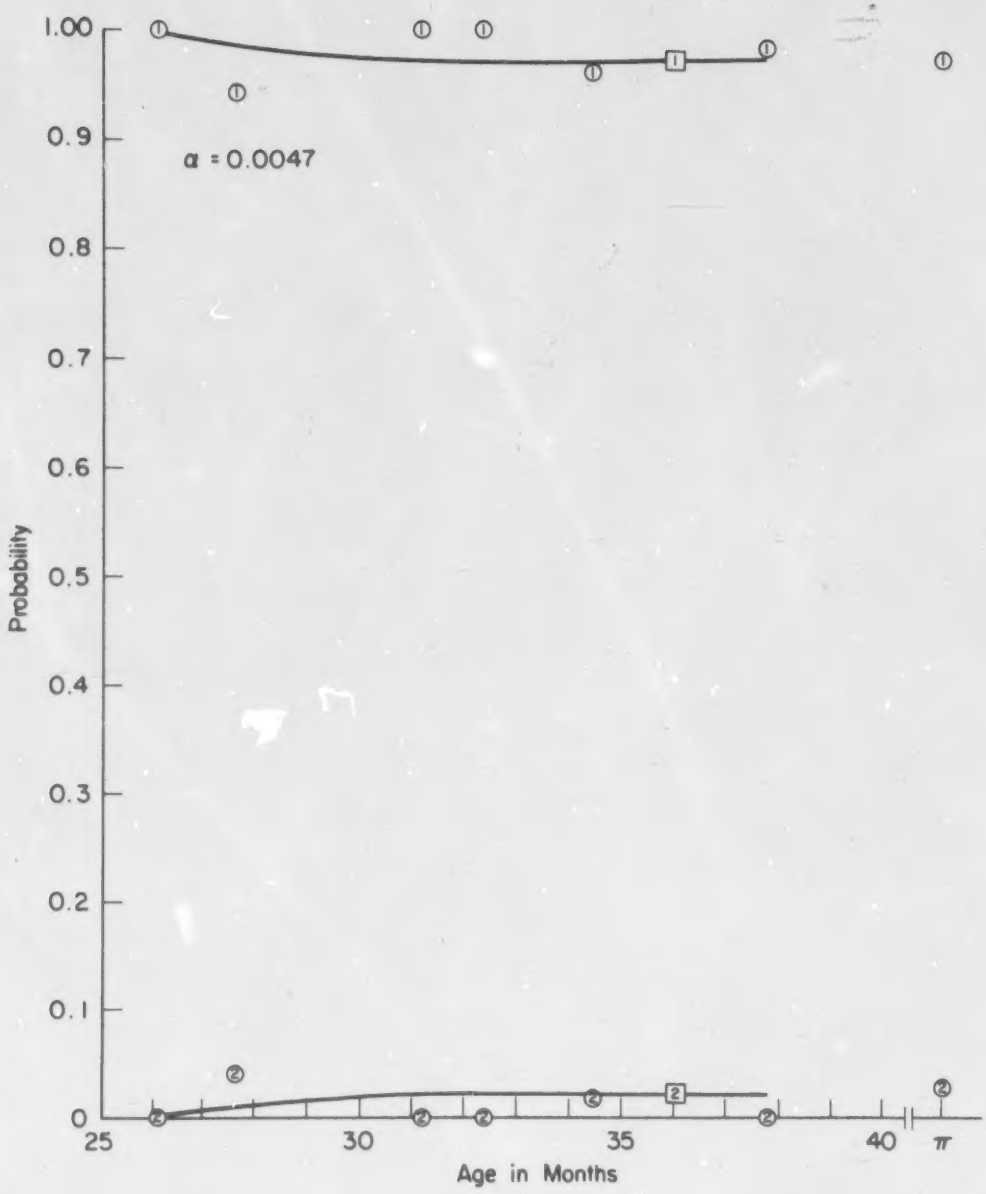


Fig. 20. Fit of incremental model for rules of Group 19, auxiliaries.

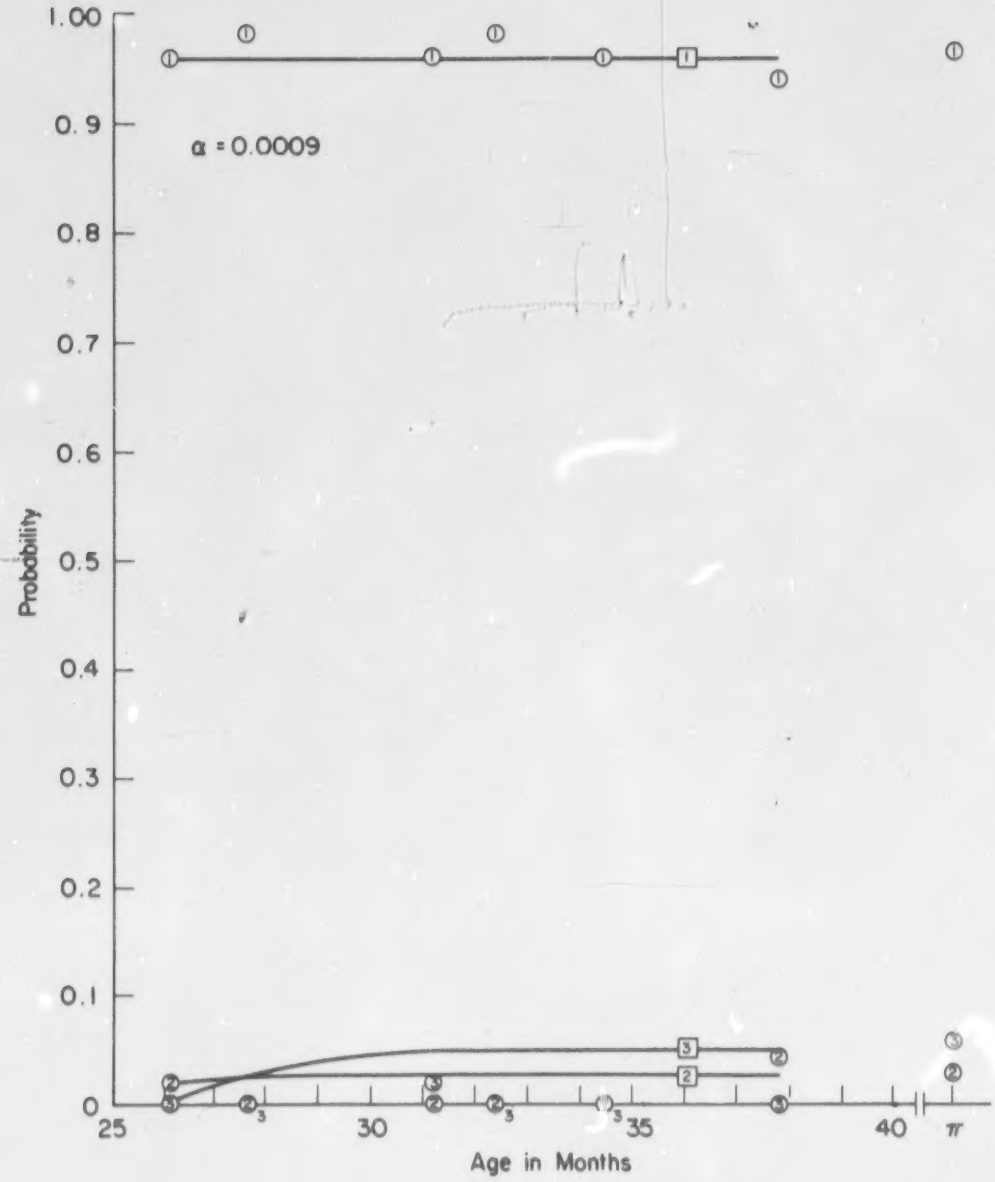


Fig. 21. Fit of incremental model for rules of Group 20, modals.

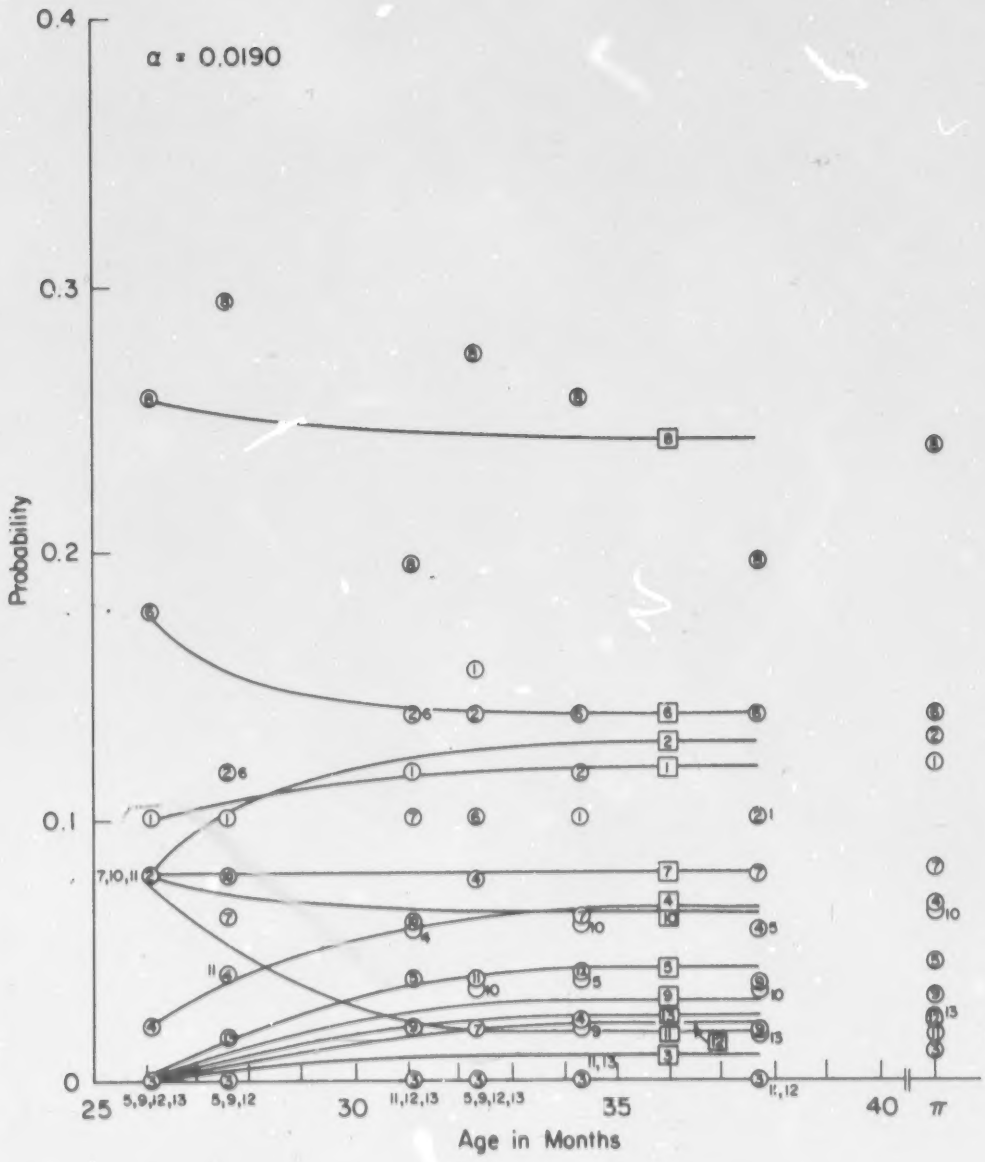


Fig. 22. Fit of incremental model for rules of Group 21, verbal groups.

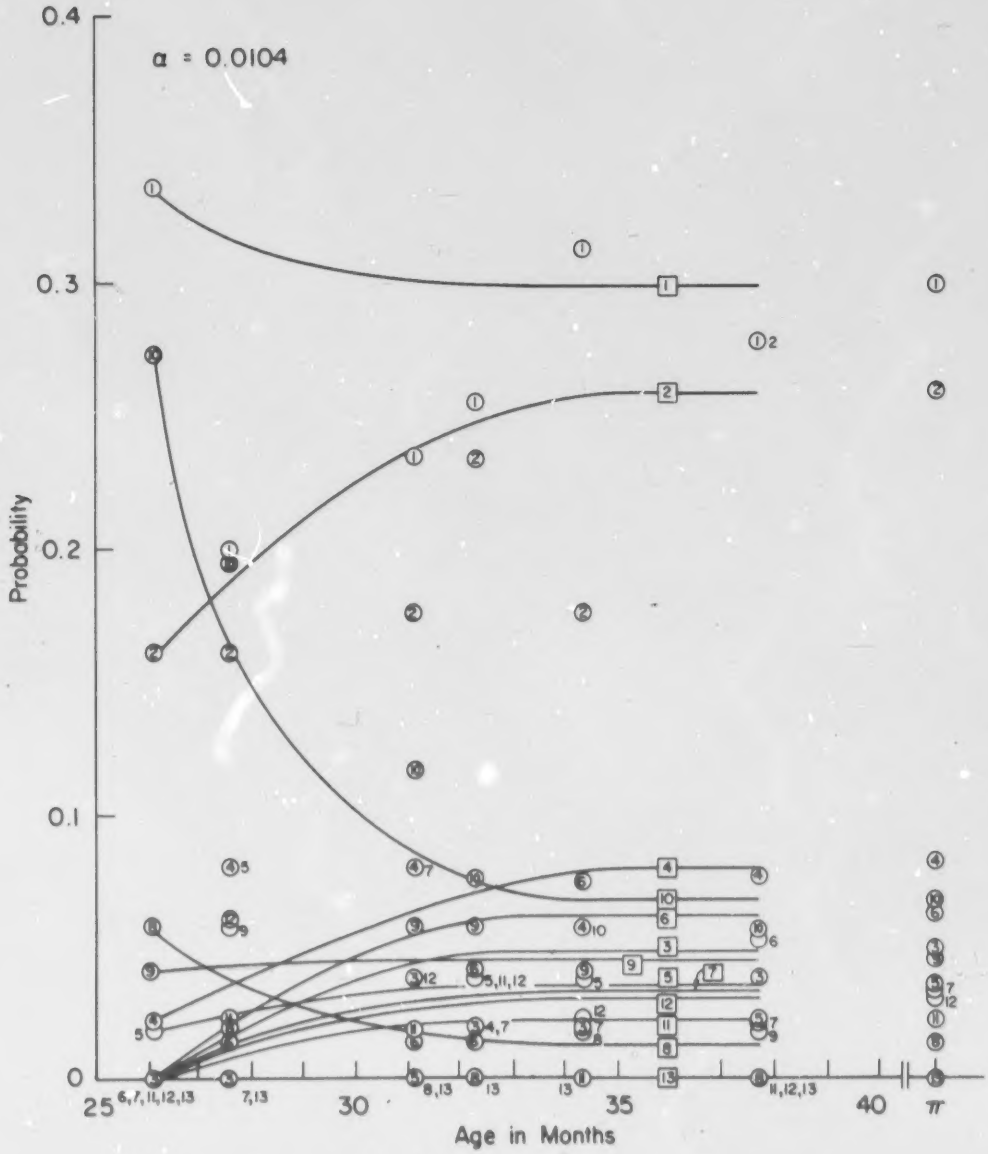


Fig. 23. Fit of incremental model for rules of Group 22, verb phrases with prepositions or personal pronouns.

The evidence seems to show that these groups do not exhibit the development of verb phrases.

Groups 23 and 25 introduce terminal symbols into verb phrases. The data and curves as shown in Figures 24 and 25 do not show any substantial developmental trends but are relatively flat.

Groups 26 and 27 (Figures 26 and 27) are low-level rules for adjective phrases, and again the data do not show any marked developmental trends.

Group 28 introduces questions. In a rough qualitative way, some of the predictions are reasonable. For example, rules 28.10, 28.11, 28.13, and 28.17 all introduce pro forma questions such as qu'est-ce c'est? The tendency to use these questions decreases sharply, and this decrease is predicted very accurately as seen in Figure 28. Other rules, such as 28.8, show something of an increase, but not as evenly as predicted.

Group 29 contains two rules that introduce the expressions un peu and un petit peu. These rules have little linguistic interest, and the data and theoretical curves are very poorly matched as seen in Figure 29.

An overall observation about the accuracy of the fits of the incremental model is that it fits best the data for the highest level groups of rules. One reason for this could be that there is more data to support the predictions made on these rules. Another reason may be that it was clearer what linguistic considerations were relevant to the top-level rules as opposed to the lower-level rules.

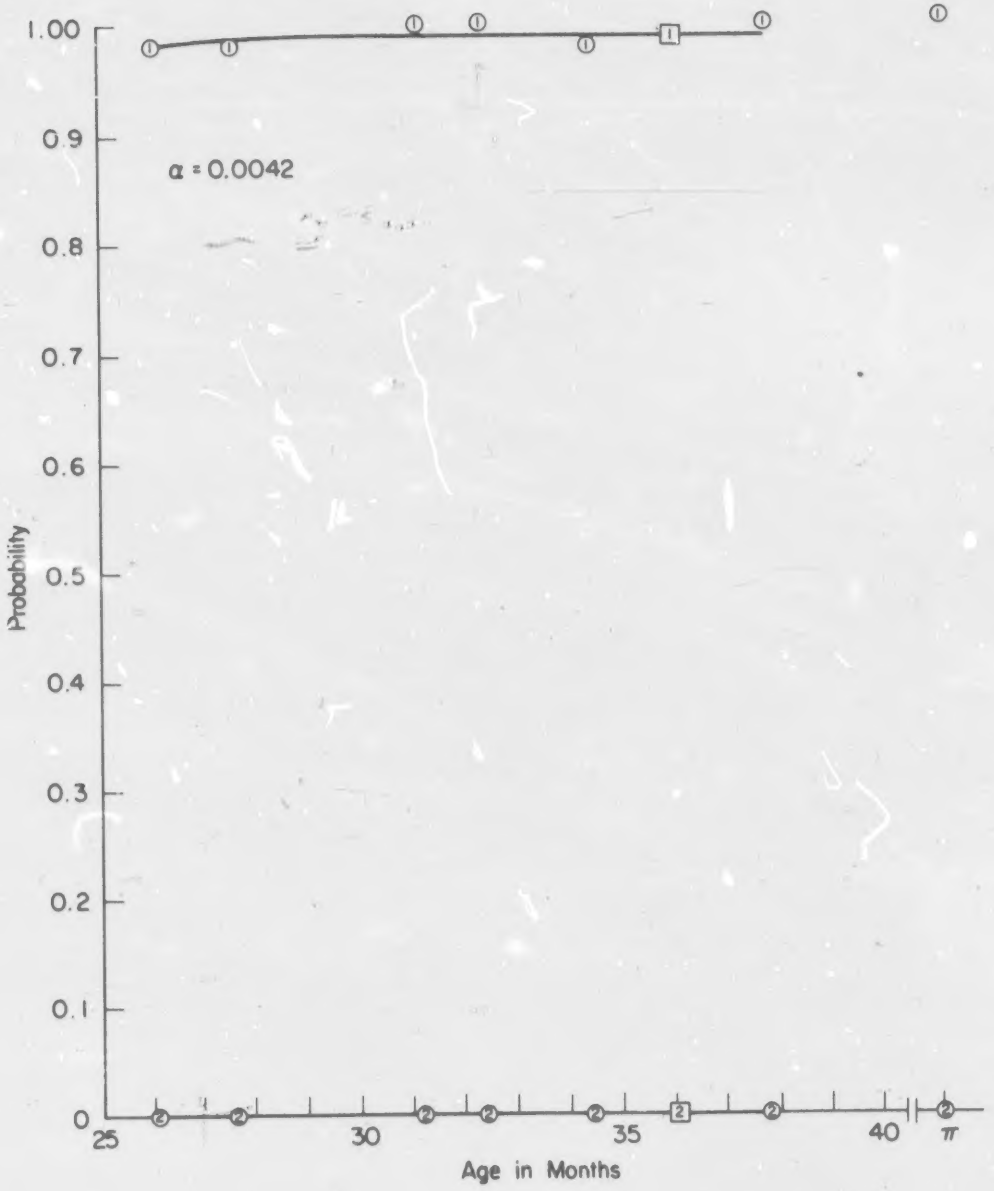


Fig. 24. Fit of incremental model for rules of Group 23, intransitive and impersonal verbs.

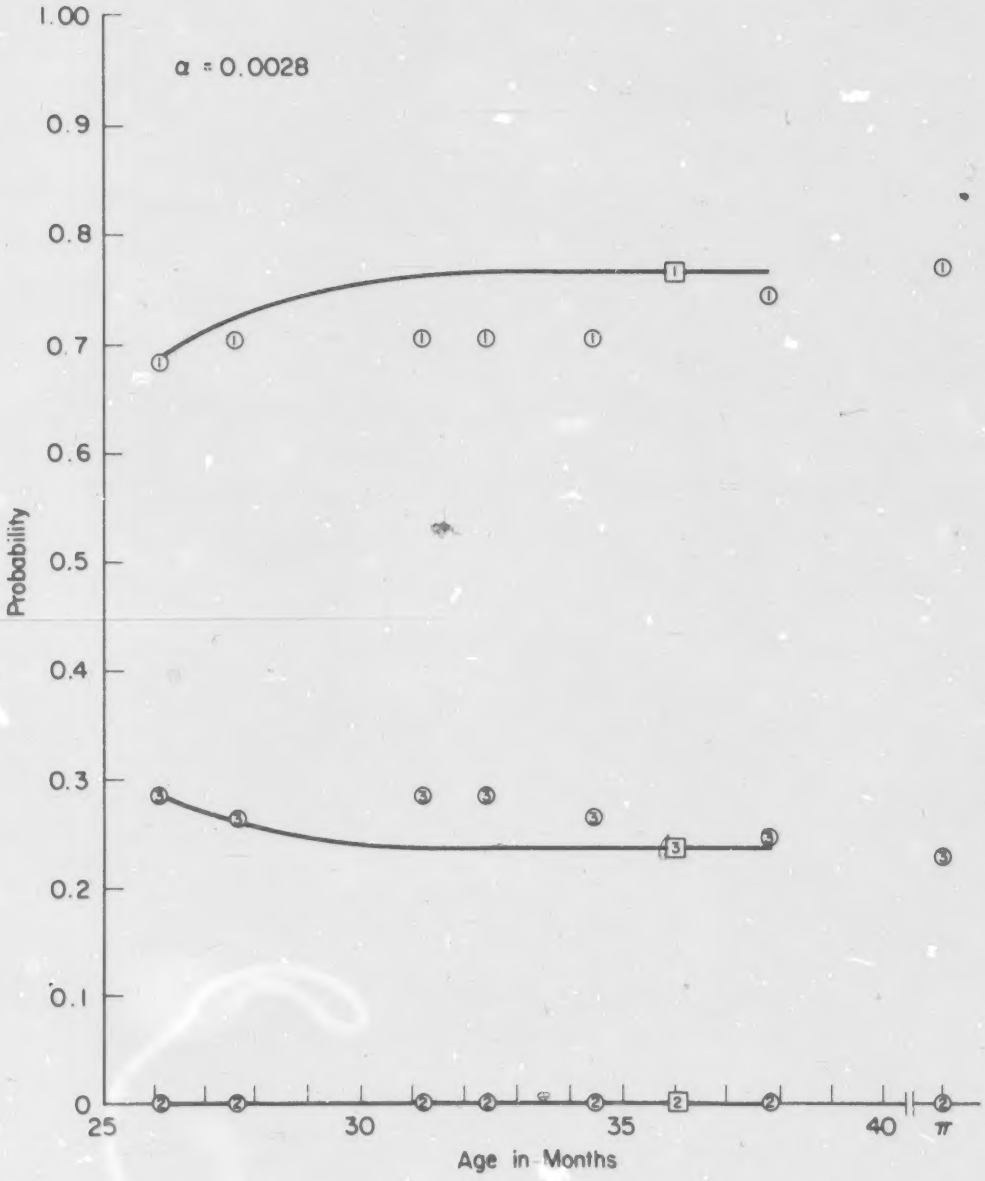


Fig. 25. Fit of incremental model for rules of Group 25, transitive and pronominal verb

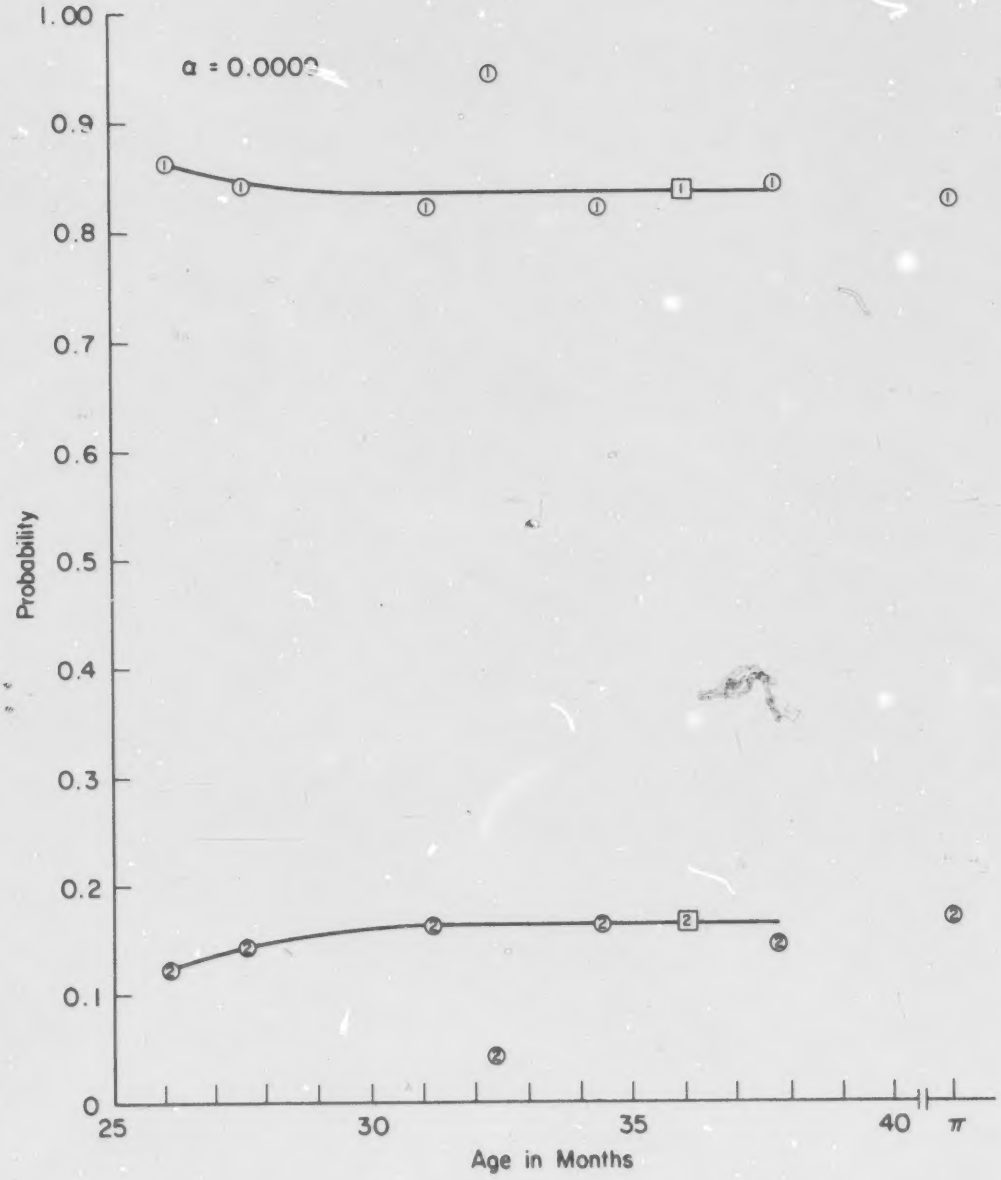


Fig. 26. Fit of incremental model for rules of Group 26, predicate adjectives.

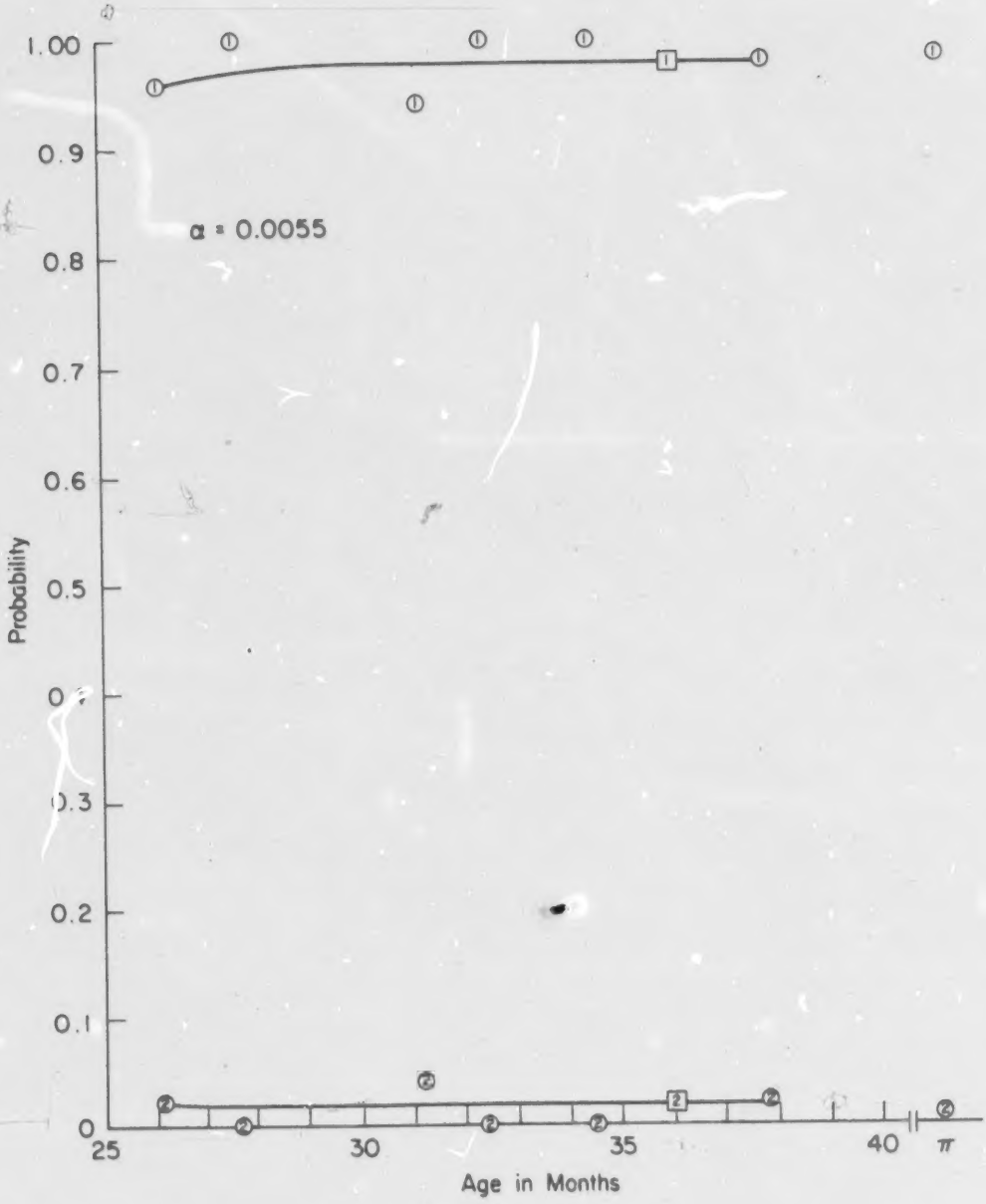


Fig. 27. Fit of incremental model for rules of Group 27, adjectives.

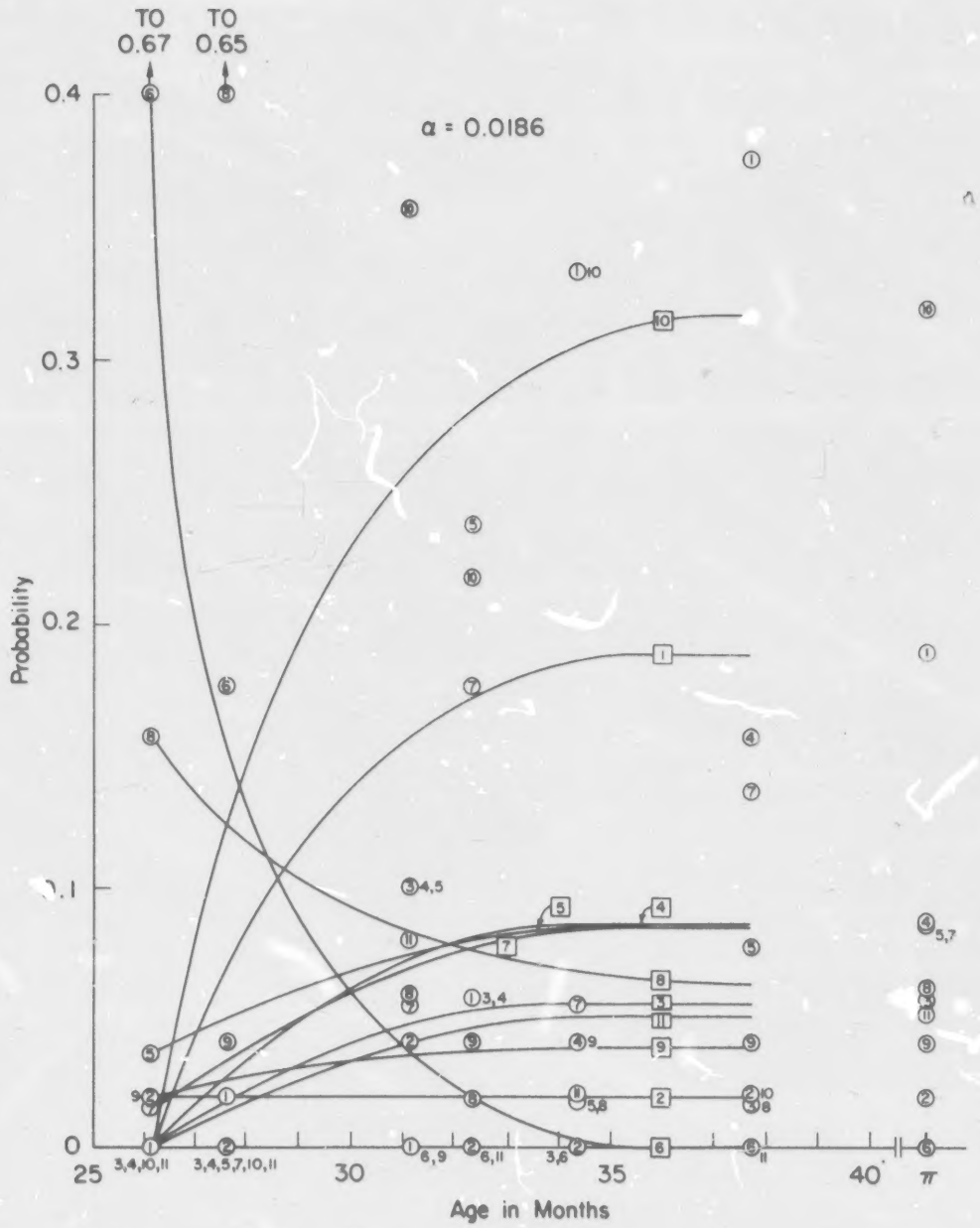


Fig. 28. Fit of incremental model for rules of Group 28, questions.

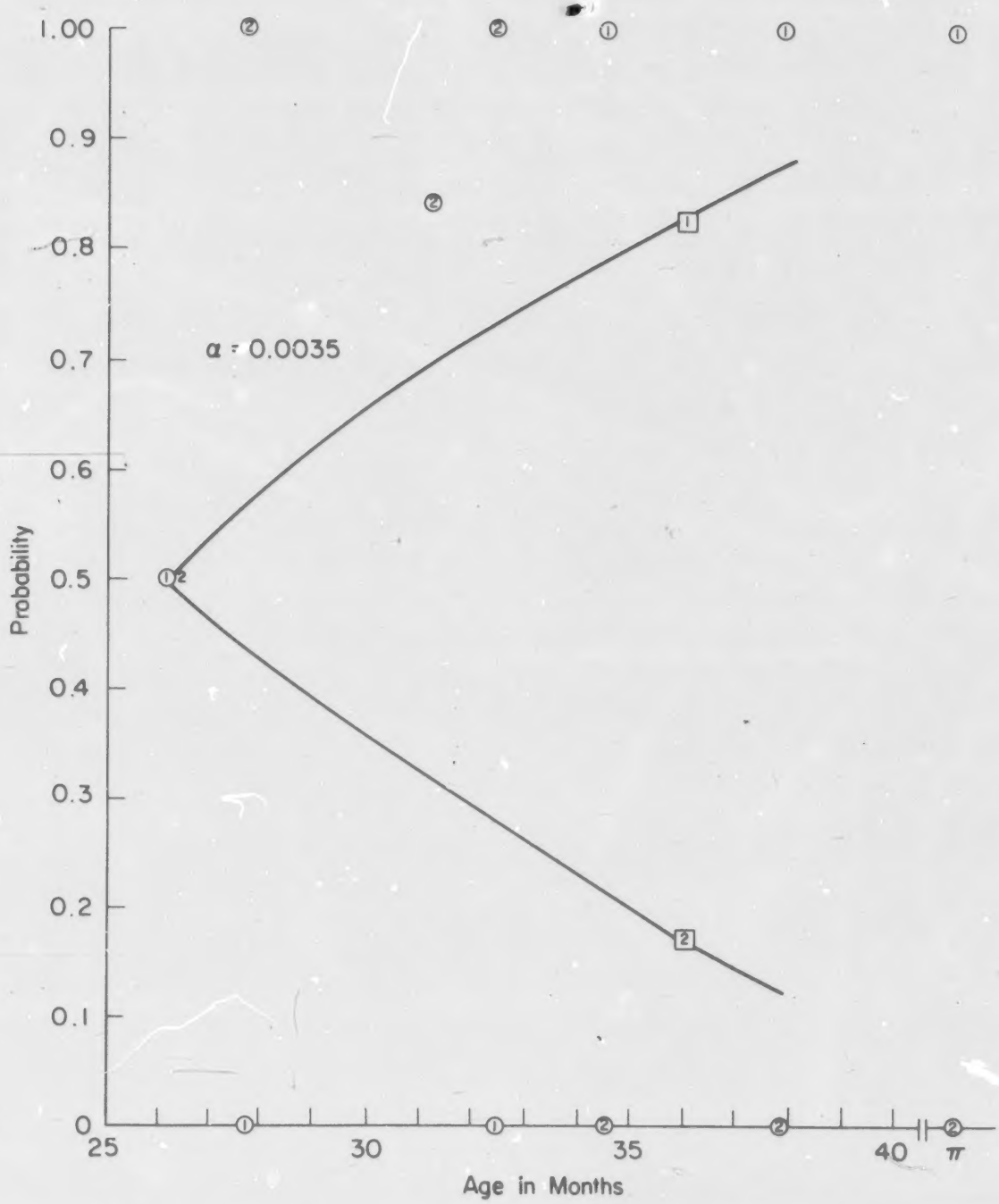


Fig. 29. Fit of incremental model for rules of Group 29, adverbs of quantity.

We have not shown here the estimates for the developmental parameter α of each group, because there is no simple way of interpreting the psychological meaning of the rate expressed by α. Due to the fact that in many cases there is very little change between the initial and final frequency of usage of the rules of a given class, the value of α can change in a relatively insensitive way over a quite wide range. In other words, with respect to α the least-squares function is very flat around the minimum. Also, the numerical value of α is determined by the units of measurement for time we have used, namely, days, and a direct psychological interpretation of these units is also not particularly interesting. Finally, it should be observed that we could have improved the fit of the theoretical curves to the data by estimating simultaneously the parameters p_r, π_r, and α for each group rather than estimating p_r first and then making an estimate of π_r and α. As indicated already, however, we do not believe that this would have led to a markedly better increase in the fit, and the computations were considerably simpler but still extensive for the approach we chose.

What we think is important about our results is this. Systematic developmental trends for the use of individual grammatical rules are easily identified in many cases and a good theoretical account of them can be given to a first approximation by a probabilistic model of an incremental character.

Appendix I

Incorrect Grammatical Analyses

A detailed analysis of the rules by which the utterances were recognized has enabled us to see that some of them were analyzed incorrectly. We present below the cases in which the grammatical analysis is not satisfactory. As will be seen, in many cases the reason the utterances are not analyzed correctly is either because Philippe was repeating the same word, or because the subject of the verb was omitted.

We present the sentence type, followed by several utterances, and by the grammatical rule which is responsible for the recognition of these utterances.

TABLE VIII

Incorrect Grammatical Analyses

VN DN NC
fume la maîtresse
brûler le feu
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 8 tokens out of 31)

VN NP
fume Philippe
tiens Madeleine (NP is a vocative)
travaille Papa (probably for Papa travaille)
ouvre Madeleine (NP is a vocative)
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 15 tokens out of 16)

VT NP
boire Madeleine
écoute Philippe (NP is a vocative)
attends Papa (NP is either a vocative or an object)

```
ramasse Madeleine (NP is a vocative)
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 9 tokens out of 14)

NC NP
docteur Rogalski
(uses rule 3.3 SN2 -> SN NP)
(incorrect analysis for 1 token out of 10)

EP NC EP DN NC
au jardin dans l'eau
(uses rule 12.12 SN*1 -> EP NP* EP NP*)
(incorrect analysis for 1 token out of 7)

EP NC EP NC
des bras des mains
(uses rule 12.12 SN*1 -> EP NP* EP NP*)
(incorrect analysis for 1 token out of 6)

NP NP
mamie mamie
(uses rule S2 -> SN NP)
(incorrect analysis for 4 tokens out of 6)

VN NC
tiens pompier
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 6 tokens out of 6)

VN IN NC
tiens un canard
tiens un couteau
(uses rule 21.25 GV -> V SNP)
(in these cases, "tiens" could be paraphrased by "here is")
(incorrect analysis for 3 tokens out of 5)

VT PT
attends toi       (for "toi, attends")
déshabillé moi    (for "moi, je suis déshabillé")
écris moi         (for "moi, j'écris")
mettre toi        (for "toi, mets ca")
renversé moi      (for "moi, j'ai renversé quelque chose"
                   ou "moi j'ai été renversé")
(uses rule 21.37 GV -> V PT)
(incorrect analysis for 5 tokens out of 5)

VT DN NC
faisaient les enfants
font les dames
(uses rule 21.25 GV -> V SNP)
```

```
(incorrect analysis for 3 tokens out of 5)

PF VN DN NC
il brûle le pistolet
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 1 token out of 4)

DV PF PS VT NP
comment elle s'appelle Clara
comment elle s'appelle Marc
(uses rule 22.2 GV' -> PS V SNP)
(incorrect analysis for 3 tokens out of 3)

DN NC VN DV
le micro parler dedans
(uses rule 21.27 GV -> V DV)
(incorrect analysis for 1 token out of 3)

DV PF VA VT DN NC
oui il était habillé le chameau
voilà il est enlevé le siège
voilà il est réparé le moteur
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 3 tokens out of 3)

EP NC NP
au parc Montsouris
(uses rule 3.4 SN2 -> SN NP)
(incorrect analysis for 1 token out of 3)

PD VC AQ NP
c'est beau Clara
c'est dur Maman
(uses rule 21.5 GV -> COP SA SN)
(incorrect analysis for 3 tokens out of 3)

VT NP
fais Maman (for "que fais-tu Maman?")
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 3 tokens out of 3)


PD VC NC NC
c'est monsieur loup
(uses rule  21.15 GV -> COP SN SNS)
(incorrect analysis for 3 tokens out of 3)

EP NC NP
de monsieur Boirat
de monsieur Seguin
```

```
(uses rule 3.3 S2 -> SN NP)
(incorrect analysis for 2 tokens out of 2)

DV PF VA VT AD NC
comment elle est faite cette maison
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 2 tokens out of 2)

PF PS VT DN NC
il se case le camion
(uses rule 22.2 GV' -> PS V SNP)
(incorrect analysis for 1 token out of 2)

PF VA VN DN NC
elle est cassée la vitre
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 2 tokens out of 2)

PF VA VN DN NC EP NP
elle est arrêtée la moto de Gerard
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 2 tokens out of 2)

PF VA VT DN NC
elle est tordue la page
(uses rule 21.24 GV -> V SNP)
(incorrect analysis for 2 tokens out of 2)

VD PS VT NP
veux le mettre papa
(uses rule 22.2 GV' -> PS V SNP)
(incorrect analysis for 2 tokens out of 2)

VD VI DN NC
faire partir la colique
faire tomber les crayons
(uses rule 21.54 GV -> V' SNS)
(incorrect analysis for 2 tokens out of 2)

PD PD VC IN NC NC
ça c'est un pont de
ça c'est un camion betonneuse
(uses rule 21.15 GV -> COP SN SNS)
(incorrect analysis for 3 tokens out of 3)

PD VC NP NP
c'est papa maman
c'est Babar Babar
c'est Gérard Martine
(uses rule 21.15 GV -> COP SN SNS)
```

(incorrect analysis for 3 tokens out of 3)

PF VC incomplete sentence
ils sont
il est
(3 tokens out of 3)

PF VA VI incomplete sentence
j'ai failli
(1 token out of 3)

VN EP NP (should have a derivation such as NP
souffler dessus papa should be an apposition)
(incorrect analysis for 3 tokens out of 3)

IN AC NC
une deux trous
(uses rule 11.12 NP* -> DET NUM NC)
(2 tokens out of 2)

IN NC EP AC NC VN
une voiture avec deux portes ouvertes
(uses rule 5.2 S' -> SNS SV)
(incorrect analysis for 2 tokens out of 2)

IN NC PF VT
un jouet je veux
(uses rule 7.3 SNS -> SN* PF)
(incorrect analysis for 2 tokens out of 2)

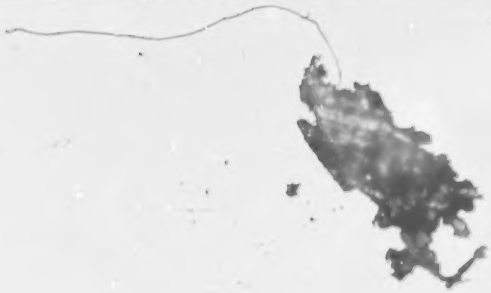
IN NC VT DN NC
une bêtise toucher la pipe
une pelle ramasser les feuilles
(uses rule 5.2 S' -> SNS SV)
(incorrect analysis for 2 tokens out of 2)

IN NC PF VT
un sucre tu manges
(uses rule 7.3 SNS -> SN* PF)
(incorrect analysis for 2 tokens out of 2)

KO DV PD PD VC should be KO PR
ah que ce c'est
(no rule KO PR PD SV)
(2 tokens out of 2)

PD VC EP NC EP NC
c'est des fruits des fruits
(rule 10.2 SN** -> EP NP* EP NP*)
(incorrect analysis for 1 token out of 2)

```
PD VD VT
celui-là faut couper
(uses rule 5.2 S' -> SNS SV)
(incorrect analysis for 1 token out of 2)

PF VI EP PI DN NC
il revient tout à l'heure
(uses rule 21.54 GV -> V' SNS)
(incorrect analysis for 2 tokens out of 2)

PF VD DG VT PT
je veux pas boire moi
(uses rule 21.37 GV -> V PT)
(incorrect analysis for 2 tokens out of 2)

PF VT NC DV
il fait camion dehors
(uses rule 21.26 GV -> V SNP DV)
(incorrect analysis for 2 tokens out of 2)

PI EP IN AQ NC
tout d'un seul coup
(uses rule 12.13 SN*1 -> NP* EP NP*)
(incorrect analysis for 2 tokens out of 2)

PR LC PF VT AD NC
qu'est-ce ils font ces gens
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 2 tokens out of 2)

PR LC PY VT EP AQ NP
qu'est-ce y a de beau Clara
(uses rule 11.23 NP* -> ADJP NP)
(incorrect analysis for 2 tokens out of 2)

CS PF PS VT NP
s'il te plaît Papa
(uses rule 22.2 GV' -> PS V SNP)
(incorrect analysis for 2 tokens out of 2)

VA VN DN NC
est cassé les fleurs
est sorti les vaches
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 2 tokens out of 2)

VA VT DN NC
est mouillée la voiture
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 2 tokens out of 2)
```



```
VT PI EP DN NC
attends tout à l'heure
(uses rule 21.25 GV -> V SNP)
(incorrect analysis for 2 tokens out of 2)

VA VT DN NC CC DN NC AQ
a mettre le gilet et les chaussons bleus
(this "a" stands probably for "je vais" or "il faut")
(incorrect analysis for 2 tokens out of 2)


FREQUENCY 1

AC AQ NC PF VT
deux petite disques je mets
(uses rule 7.3 SNS -> SN* PF)

AC AQ NC PF PS VT
deux petites filles elles s'appellent
(uses rule 7.3 SNS -> SN* PF)

AC NC PF VA VT
trois heures t'avais mis
(uses rule 7.3 SNS -> SN* PF)

AD NC PF VT DG
ces gens je sais pas
(uses rule 7.3 SNS -> SN* PF)

AD NC AD PD VC NP AD NC                    repetition
cette dame-là c'est Madeleine cette dame
(uses rule 25.15 GV -> COP SN SNS)

AD NC AD PF VA VN AD NC                    repetition
cette roue-là elle est cassée cette roue
(uses rule 21.25 GV -> V SNP)

AD NC AD PF PS VT DG AD NC                 repetition
ce chien-là je le connais pas ce chien
(uses rule 7.3 SNS -> SN* PF)

AO NC DV PF PS VT NC
ton stylo-là j'en ai besoin
(uses rule 7.9 SNS -> SN* DV PF)

AO NC NP
mon ami Pierrot
(uses rule 3.4 S2 -> SN NP)

AO NC PF PS VT AO NC                repetition
```

```
ta fourrure tu la mets ta fourrure
(uses rule 7.3 SNS -> SN* PF)

AO NC PF PS VT EP AO NC
son moteur il le laisse dans sa tête
(uses rule 7.3 SNS -> SN* PF)

AQ AQ NC
cochon sales mains
(uses rule 11.11 NP* -> ADJP NC)

AT NC PD VC PD
quelle couleur c'est ça
(uses rule 7.7 SNS -> SN* PD)

AT NC PF PS VA VT DN NC
quelle couleur tu l'as achetée ta montre
(uses rule 7.3 SNS -> SN* PF)

AT NC VD VN
quel disque faut prendre
(uses rule 5.2 S' -> SNS SV)

AT NP NP
quelle Madame Rogalski
(uses rule 3.4 S2 -> SN NP)

CC DN NC PF VI                    should be AQ instead of NC
et le bleu il marche


CC DN NC EP EP DN NC              stuttering
et les mains de au le monsieur
(uses rule 12.4 SN*1 -> NP* EP EP NP*)

CC DN NC EP EP NC                 stuttering
et les bouteilles de d'eau
(uses rule 12.4 SN*1 -> NP* EP EP NP*)

CC DV DN AQ NC PF VC EP IN NC EP IN NC
et la la petite souris elle est sur une casserole sur une poubelle
(uses rule 10.2 SN** -> EP NP* EP NP*)

CC DV IN NC PF VA VT
et là une pomme elle a mis
(uses rule 7.3 SNS -> SN* PF)

CC DV DV DN NC PF PS VT
et voilà que le matin il la mange
(uses rule 4.16 S3 -> CC DV DV S' and 7.3 SNS -> SN* PF)
```

```
CC IN NC EP DN NC VN
et une voiture avec le coffre ouvert
(uses rule 5.2 S' -> SNS SV)

CC NP PF VA VN PD EP NP
et Madeleine il a fondu celui de Madeleine
(uses rule 7.3 SNS -> SN* PF and 21.25 GV -> V SNP SN**)

CC PD EP AQ NC
et ça des petites lumières
(uses rule 12.10 SN*1 -> pd ep NP*)

CC PD VC PI EP NC PD              repetition
et ça c'est tous des camions ça
(uses rule 21.15 GV -> COP SN SNS)

CC PD VA VT AQ NC EP NC EP NC
et celui-là a mis ses oreilles en bas du chapeau
(uses rule 21.35 GV -> V SNP SN**)

DG PD VC DN NC CS PI VD VI        should be PR
non c'est le pont qu'on fait tomber
(there should be a rule  S' -> SNS SV PR SNS SV)


CC DV NC VT DG EP NC EP NC
et en bas a pas de peinture en bas
(uses rule 4.10 S3 -> CC DV S')

CC PD VC AO NC CS PF VA VT LC          should be PR instead of CS
mais c'est ton fou que j'ai mis la-bas

CC PF                    incomplete
mais je

DV PF PS VT NP
pas bleu rouge

DG DG PI EP NC
non pas tout de suite
(uses rule 12.3 SN*1 -> NP* EP NP*)

DG DG PI EP VT
non pas tout à fait
(uses rule 22.16 GV' -> EP V)

DG DG VN                         incomplete
non pas réfléchis

DG PS PS VT DG                   stuttering
```

```
non le l'accroche pas
(uses rule 22.13 GV' -> PS PS V DG)

DG DN NC PS NC                          should be DG DN NC DN NC
pas le front le peigne
(uses rule 22.1 GV' -> PS V)

DG DN NC PF VI
non le tabouret je veux
(uses rule 7.3 SNS -> SN* PF)

DG PD VC AQ PI                          should be PF
non c'est pas triste moi

DG PD VC DN NC CS PI VD VI              should be PR instead of CS
non c'est le sable qu'on fait tomber

DG PD VN DN NC
non ça tourne les trains
(uses rule 21.25 GV -> V SNP)

DG PF VI DG AQ NP
non il coud pas mon papa
(uses rule 21.29 GV -> V DG SNP)


DG PF VI PI
non je reviens nous
(uses rule 21.54 GV -> V' SNS)

DG PF VC DG DV NC                       should be AQ
non elle est pas toujours malade

DG VI DV                                incomplete
pas mis comme

DG VI DG EP NC EP DN NC DN NC           repetition
non a pas de manège à l'ecole à l'ecole
(uses rule 21.30 GV -> V DG SNP SN**)

DN AQ NC PF VD VN DN NC DN AQ NC        repetition
le petit garçon il veut chasser le papillon le petit garçon
(uses rule 21.24 GV -> V SNP SNS)

DN AQ NC VI IN NC CS PF PF VN   should be  DV PT PF VN
la petite fille a un magnétophone comme toi tu vois
(uses rule 5.3 S' -> SNS SV CS SNS SV and
(rule 7.4 SNS -> PF PF)

DN NC PS VI DN
```

les vis les mettre là
(uses rule 3.2 S' -> SNS SV)

DN NC AQ PF VA VT DN NC AQ DN NC (le feu is analyzed as a subject)
le feu rouge il a oublié le feu rouge le conduiteur
(uses rule 21.34 GV -> V SNP SNS)

DN NC EP NP PF VC EP DN NC NP
le syndicat de maman il est à la rue Montbrun
(uses rule 21.15 GV -> COP SN SNS)

DN NC EP NP NP
la voiture à papa maman
(uses rule 3.4 S2 -> SN NP)

DN NC EP VT CS DN NC VD VI DN NC
le sifflet pour dire que le train va partir la valise
(uses rule 21.54 GV -> V' SNS)

DN NC PD VT EP NC DN NC repetition
les fleurs ça a des mains les fleurs
(uses rule 21.34 GV -> V SNP SNS)

DN NC PF PS VA VT DN NC repetition
les croissants tu les as emportés les croissants
(uses rule 7.3 SNS -> SN* PF)

DN NC PF PS VA VT NP
les croissants tu les as emportés Madeleine
(uses rule 7.3 SNS -> SN* PF and
(rule 21.25 GV -> V SNP)

DN NC PF VC EP DN NC DN NC repetition
les gens ils sont à l'endroit les gens
(uses rule 21.15 GV -> COP SN SNS)

DN NC PF VA VN EP NC PD
le chien il est passé en dessous ça
(uses rule 21.34 GV -> V SNP SNS)

DN NC PF VA VT DN NC repetition
le toboggan il est tordu le toboggan
(uses rule 21.25 GV -> V SNP)

DN NC PF VT EP AQ NC DN NC repetition
les chiens ils ont des petites maisons les chiens
(uses rule 21.34 GV -> V SNP SNS)

DN NC PF VT AC NC DN NC repetition
la dame elle a deux bracelets la dame

(uses rule 21.34 GV -> V SNP SNS)

DN NC PF VT IN NC DN NC repetition
la maîtresse elle a un stylo la maîtresse
(uses rule 21.24 GV -> V SNP SNS)

DN NC DV PF PS VT
le matin qu'il la mange
(uses rule 7.9 SNS -> SN* DV PF)

DN NC EP NC
la fumée par terre
(uses rule 12.3 SN*1 -> NP* EP NP*)

DN NC VA VT DV
le doigt a met dedans
(uses rule 5.2 S' -> SNS SV)

DN NC VC incomplete
les garçons étais

DN NC VT
la poussière balaie
(uses 5.2 rule S' -> SNS SV)

DN NC VT EP DN NC
la casserole mettre à la cuisine
(uses rule 5.2 S' -> SNS SV)

DN NC EP EP DN NC stuttering
la fumée du de la locomotive
(uses rule 12.4 SN* -> NP* EP EP NP*)

DN PC EP NC
la mienne de raquette
(uses rule 12.3 SN*1 -> NP* EP NP*)

DV DN NC PF VT EP NC DN NC
comme le camion il faisait du bruit l'avion
(uses rule 7.3 SNS -> SN* PF)

DV DV PF VC LC should be DV CS
mieux qu'ils restent là-dedans
(uses rule 4.14 S3 -> DV DV S')

DV IN NC PF VA VN
comme un crayon il est roulé
(uses rule 7.3 SNS -> SN* PF)

```
DV PD PD VC DN NC
comme ça c'était les bateaux
(uses rule 4.3 S3 -> DV S')

DV PD PD VN
comme ça ça tourne
(uses rule 4.3 S3 -> DV S')

DV PD PD VN DM AQ
comme ça ça tourne tout seul
(uses rule 4.3 S3 -> DV S')

DV PD PD VD VI DN NC
comme ça ça fait marcher le commerce
(uses rule 4.3 S3 -> DV S')

DV PD VI DN NC
comment ça fait les moutons?
(uses rule 21.25 GV -> V SNP)

DV PF PS VN DN NC
quand elle s'arrêtera l'aiguille?
(uses rule 21.24 GV -> V SNP SNS)

DV PF PS VI DN AQ NC EP NP
comment elle s'appelle la petite fille de Madeleine?
(uses rule 22.2 GV' -> PS V SNP)

DV PF PS VI NP
comment tu t'appelles Madeleine?
(uses rule 22.2 GV' -> PS V SNP)

DV PF PS VI AD NC AD
comment ils s'appellent ces enfants-là?
(uses rule 22.2 GV' -> PS V SNP)

DV PF VA DG VI NC AD NC AD
pourquoi il a pas mis son ce manteau-là?
(uses rule 21.24 GV -> SNP SNS)

DV PF VA VI AD NC
comment il est fait ce poisson?
(uses rule 21.25 GV -> V SNP)

DV PF VN CS PF VD EP VI EP NC EP DN NC
comment tu ouvres comment tu fais pour mettre du feu dans la cigarette?
(uses rule 4.3 S3 -> DV S' and 5.3 S' -> SNS SV CS SNS SV)

DV PF VN DG DN NC
pourquoi elle tient pas la moto?
```

(uses rule 21.25 GV -> V SNP)

DV PF VN DN NC
pourquoi il pleure le bébé?
(uses rule 21.25 GV -> V SNP)

DV PF VT NP
comment tu travailles Madeleine?
(uses rule 21.25 GV -> V SNP)

DV PF VT NP
comment tu dessines Madeleine?
(uses rule 21.25 GV -> V SNP)

DV PS VT DV
là le peigne là
(there is no rule DV DET NC DV)
(uses rule 22.4 GV' -> PS V DV)

DV DV PF VA VI DV repetition
alors là ils sont partis là

DV EP EP NC PI VT EP AQ NC stuttering
là dans au magasin on acheté des petits chiens
(uses rule 5.7 S' -> SNS** SNS SV)

DV DN AQ NC VT DV repetition
là un petit trou fais là
(uses rule 5.2 S' -> SNS SV)

DV IN NC PF VT DG
dedans une chose je sais pas
(uses rule 7.3 SNS-> SN* PF)

DV IN NC DV PF VA VT should be PR instead of DV
là un chien qu'ils ont laisse là
(uses rule 7.9 SNS -> SN* DV PF)
(no rule S' -> SN* PR SNS SV)

DV PD VC EP VT EP EP AQ NC should be NC instead of VT
là c'est des couverts pour des petits enfants
(no rule SN** -> EP NP* EP EP NP*)
(uses rule 21.23 GV -> COP GV')

DV PD VC IN NC EP EP NC EP VN PR VN should be EP NC instead of EP VN
là c'est un autocar avec du fil de courant qui passe
(no rule SN* -> NP* EP EP NP* EP NP*
(uses rule 5.5 S' -> SNS SV PR SV)

DV PF VT EP PI AQ NC should be AI instead of PI

oui il coupe avec plusieurs gros trucs
(no rule NP* -> AI AQ NC)
(uses rule 21.34 GV -> V SNP SNS)

DV PF VC AQ NC
dedans il est blanc bateau
(uses rule 21.5 GV -> COP SA SN)

DV PF VA VN NP 2 derivations should be necessary
voila j'ai fini Ginette
(no rule GV -> V SNS)
(uses rule 21.25 GV -> V SNP)

DV AQ NC should be DM
trop petit chausson

DV PF VC AO NC AQ NP
où il est ton ciré gentil papa
(uses rule 21.15 GV -> COP SN SNS)

DV PF VC DN NC should be DN AQ
où il est le bleu

DV PF VC DN NC AQ NC
où elle est la voiture grosse voiture
(uses rule 21.15 GV -> COP SN SNS)

DV PF VC PD DN NP EP AD NC stuttering
où il est ce le gronomme de ce monsieur?
(uses rule 21.15 GV -> COP SN SNS)

DV PF VC PS VI DN AI NC
où il est le su l'autre sucre? stuttering
(uses rule 21.13 GV -> COP GV' and 22.2 GV' -> PS V SNP)

DV PF VC DN NC AI NC
où elle est la roue l'autre roue?
(uses rule 21.15 GV -> COP SN SNS)

DV PF VC NP EP DN NC
où elle est Mamie dans le salon
(uses rule 9.3 SN* -> NP* EP NP*)

DV PF VC NC EP DN NC DV
où elle est Mamie dans le salon là
(uses rule 21.4 GV -> COP SN and rule 9.3 SN* -> NP* EP NP*)

DV PF VD PS VI DV repetition
après il va les ramasser après

```
DV PI PF VN DV AD NC
là on il travaille aussi ce monsieur
(uses rule 7.3 SNS -> SN* PF and
(rule 21.28 GV -> V DV SNP)

DV PI VN PR PR PD VC CS PF VC AQ NP      stuttering
après on croit que que c'est que elle est vieille Limoges
(uses rule 21.52 GV -> V SNP PR S')

DV VA VT NP      incomplete
pourquoi a enlevé Madeleine?

DV VT AO NC AQ NC
voilà écrire ton nom petit cochon
(uses rule 21.34 GV -> V SNP SNS)

DV VT DG DV
en a pas ici
(no rule PS V DG DV)

DV VT EP NC EP AO NC EP PT
pourquoi a des trous à tes chaussons à toi?
(uses rule 21.35 GV -> V SNP SN**)


EP DN NC PF PS VT DN NC EP DN NC
dans la voiture je le mets le tube dans la voiture
(uses rule 22.2 GV' -> PS V SNP)

EP PS VN DN NC                    should be EP DN NC DN NC
sur la figure la poupée
(uses rule 22.19 GV' -> EP PS V SNP)

EP EP AT NC PF VT                 stuttering
à à quelle heure tu veux
(uses rule 5.7 S' -> SN** SNS SV)

EP EP IN NC                       stuttering
avec du un pinceau
(uses rule 12.16 SN*1 -> EP EP NP*)

EP IN PI NC VT                    stuttering
des une autre lunette accrochée
(uses rule 5.7 S' -> SN** SNS SV)

EP NC EP PS VN DN NC              should be EP NC EP DN NC DN NC
du crayon sur la figure la poupée
(uses rule 5.8 S' -> SN** GV')

EP AI NC IN NC EP NC PF VT
```

avec plusieurs machines une scie à moteur il coupe
(rule 5.7 S' -> SN** SNS SV and
(rule 7.3 SNS -> SN* PF)

EP PR PD VC EP VT PR PD
pour quoi c'est pour faire quoi ça
(uses rule 5.7 S' -> SN** SNS SV)

EP PR PD VT IN NC
à quoi ça sert une fronde
(uses rule 21.25 GV -> V SNP)

EP PR PF VN NP
à quoi il joue Gérard?
(uses rule 21 25 GV -> V SNP)

EP PR PF VN DN NC
par quoi il sort le feu?
sur quoi elle roule la voiture?
(uses rule 21.25 GV -> V SNP)

EP VN NP
pour parler maman

EP VI NC should be EP VD VI
pour faire rire
(no rule EP VD VI)
(uses rule 22.17 GV' -> EP V SNP)

IN AQ NC EP NC PY VT IN NC
un petit bout de roue y a une vis
(uses rule 5.2 S' -> SNS SV and 21.64 GV -> PY V SNP)

IN AQ NC PF VT
une petite cloche elle a
(uses rule 7.3 SNS -> SN* PF)

IN AQ NC PY VI EP AD NC
un petit train y a sur ce livre
(uses rule 7.3 SNS -> SN* PF)

IN AQ NC EP NC EP VT EP EP DN NC
un petit bout de pain pour mettre dans dans le chocolat
(uses rule 8.2 SNP -> EP EP NP*)

IN AQ NC PF VT
une carte postale il envoie
(uses rule 7.3 SNS -> SN* PF)

IN NC PD VC DN NC DN NC
un chariot c'est les gens les chariots
(uses rule 21.34 GV -> V SNP SNS)

IN NC PF VT EP NP
un cendrier je veux pour Madeleine
(uses rule 7.3 SNS -> SN* PF)

IN NC PF VN
une voiture elle roule
(uses rule 7.3 SNS -> SN* PF)

IN NC PF VT AQ NP
un chateau elle a ma Maman
(uses rule 7.3 SNS -> SN* PF)

IN NC DV PF VA VT should be PR instead of DV
un chien qu'ils ont laissé
(uses rule 7.9 SNS -> SN* DV PF)

IN NC VD VT DV
un âne faut faire là
(uses rule 5.2 S' -> SNS SV)

IN NV VN EP NC
un monsieur vu au cinéma
(uses rule 5.2 S' -> SNS SV)

IN NC VI
un trou arrivé

IN NC VN IN NC
une lampe manque un fil
(uses rule 5.2 S' -> SNS SV)

IN NC VT DN NC EP DN NC
une belle ramasser les feuilles dans le rateau
(rule 5.2 S' -> SNS SV)

KO NC should be AQ
ah idiot
(no rule KO AQ)

KO NC should be KO KO
ah bon
(no rule KO AQ)

KO DV PF VI DG
attention qu'elle tombe pas
(uses rule 4.12 S3 -> KO DV S')

KO DV PS VN EP DN NC should be DN NC instead of PS VN
oh encore la fumée dans la musique
(no rule S2 -> KO DV SN*1)

KO EP DN NC EP EP NC
oh dans la chaise dans des bancs
(uses rule 12.15 SN*1 -> EP NP* EP EP NP*)

KC EP NC EP NC stuttering
oh des bols des bols
(uses rule 12.12 SN*1 -> EP NP* EP NP*)

KO NC PF VD VN should be KO KO
attention attention je vais fermer
(uses rule 7.3 SNS -> SN* PF)

KO NC PF VI DG
ben dimanche je sais pas
(uses rule 7.3 SNS -> SN* PF)

KO PD VC IN NC IN NC
oh c'est un oiseau une tourterelle
(uses rule 21.15 GV -> COP SN SNS)

KO PF VA VN AO NC
oh il est roulé ton sac
(uses rule 21.25 GV -> V SNP)

KO PF VA VN DN NC
oh il est rentré la pointe
(uses rule 21.25 GV -> V SNP)

LC PF PF VI EP DN NC should be LC PF PS
parce il lui tombe de la farine
parce il lui tombe de la neige
(uses rule 7.4 SNS -> SN* PF)

NC EP NC
mal au ventre
(uses rule 12.3 SN*1 -> NP* EP NP*)

NC PD VC
losange c'est
(uses rule 7.7 SNS -> SN* PD)

NC PD VC IN NC
caisse c'est un tonneau
(uses rule 7.7 SNS -> SN* PD)

NC PF VC should be AQ

```
fou elle est
(uses rule 7.3 SNS -> SN* PF)

NC PF VC DV
bouton il est là
(uses rule 7.3 SNS -> SN* PF)

NC VD VI
chaise allez viens
(uses rule 5.2 S' -> SNS SV)

NC VN DV
micro parler dedans
(uses rule 5.2 S' -> SNS SV)

NC VN NC
pompier tiens pompier
(uses rule 5.2 S' -> SNS SV)

NC EP DN NC VT                    should be VT
bois de l'eau fais
(uses rule 5.2 S' -> SNS SV)

NP DV PF PS VT DV
Jean-Michel aussi je l'aime bien
(uses rule 7.9 SNS -> SN* DV PF)

NP PD VC DG AQ NP                 repetition
marco c'est pas possible marco
(uses rule 7.7 SNS -> SN* PD and 21.4 GV -> COP DG SA SN)

NP PF VD VT
maman je vais attaquer
(uses rule 7.3 SNS -> SN* PF)

NP PF VI AQ NC NP                 repetition
Papa il a un petit cinema papa
(uses rule 7.3 SNS -> SN* PF and rule 21.34 GV -> V SNP SNS)

NP PF VT DG
Rogalski je connais pas
(uses rule 7.3 SNS -> SN* PF)

NP PF PS VT IN NP                 repetition
Jeanine elle nen veut un Jeanine
(uses rule 21.25 GV -> V SNP and rule 11.21 NP*-> DET NP)

NP PF VC DG DV NP                 repetition
Nicole elle est pas là Nicole
(uses rule 21.19 GV -> COP DG DV SN)
```

```
NP PR VT PD
Assi-les-Moulineaux qui dit ca?
(uses rule 5.6 S' -> SNS PR SV)

NP PS PS VA VT IN NP                repetition
Minou m'en a acheté un Minou
(uses rule 21.25 GV - V SNP and rule 11.21 NP* -> DET NP)

NP VN DN NC NP                      repetition
Papa donne la main Papa
(uses rule 21.34 GV -> V SNP SNS)

NP VT DN NC NP                      repetition
Papa regarde le serpent Papa
(uses rule 21.34 GV -> V SNP SNS)

NP VT IN NC NP                      repetition
Madeleine fait un avion Madeleine
(uses rule 21.34 GV -> V SNP SNS)

PD EP NC
ça des champignons
(uses rule 12.10 SN*1 -> PD EP NP*)

PD PD VC IN AQ NC IN NC
c'est une petite fille une poupée
(uses rule 21.15 GV -> COP SN SNS)

PD PD VC IN NC PR VD VI EP DN NC        repetition
ça c'est une voiture qui va va dans la lune
(uses rule 18.8 SV -> AUX2 GV and 20.1)

PD PS VN DN NC
ça me fatigue le chocolat
(uses rule 22.2 GV' -> PS V SNP)

PD VC DG EP NC EP AQ NC
c'est pas des moteurs des petites boules
(uses rule 21.10 GV -> COP DG SN and 10.2 SN** -> EP NP* EP NP*)

PD VC DN NC AQ NC
c'est la raquette bon côté
(uses rule 21.15 GV -> COP SN SNS)

PD VC DN NP EP EP AQ NC DV
c'est le papa de des petits cochons là
(uses rule 9.5 SN* DV -> NP* EP EP NP* DV)

PD VC IN NC CS PF VA VT             should be PR instead of CS
c'est un chien qu'ils ont laissé
```

(no rule SNS SV PR SNS SV)

PD VC PD CS PI VT — should be PR instead of CS
c'est ça qu'on fait

PD VC PT CS PF VD PS PS VN — should be PR instead of CS
c'est moi que je vais te l'arrêter

PD VC PR PS VT — coding of Fascine
c'est quoi le fasciné
(uses rule 5.5 S' -> SNS SV PR SV)

PD VC DV CS PF PS VA VT AD NC — should be PR PR
c'est où que tu l'as acheté ce balai

PD VD VI DN NC
ça fait partir la fumée
(uses rule 21.54 GV -> V' SNS)

PD VD VN PI EP DN NC — tout à l'heure sould have coded as an adverb
ça va parler tout à l'heure

PD VN DG DN NC
ça vole pas les chats
(uses rule 5.2 S' -> SNS SV)

PD VN PD
ça ouvrir ça

PD VT DG NC — "bon" is analyzed as a noun
ça sent pas bon

PF PF PS VI DG EP NC
je t'en fais pas de baiser
(rule 7.4 SNS -> PF PF)

PF PS VT DN NC
elles s'amusent les tortues
(uses rule 22.2 GV' -> PS V SNP)

PF PS VT DG DN NC
elle s'ennuie pas la vache
(uses rule 22.5 GV' -> PS V DG SNP)

PF PS VT DV PI EP PT
elle m'amène quand même chez elle
(uses rule 9.6 SN* -> NP* EP NP*)

PF PS VD VN CS PF VT IN AQ NC

tu me viens voir comment je fais un petit bateau

PF PS VA VN EP EP DN NC repetition
il s'est couché sur sur le lit
(uses rule 10.5 SN** -> EP EP NP*)

PF PS VN DN NC
il se ferme le briquet
(uses rule 22.2 GV' -> PS V SNP)

PF PS VT IN DV DV should be IN DM DM
tu en veux un peu plus
(uses rule 4.15 S3 -> S' DV DV)

PF PS VA VT EP DN NC EP NC DN NC
je l'ai mis dans le coffre du camion le triangle
(uses rule 21.34 PS VA GV -> PS VA V SNP SNS)

PF VA DV VN DN NC EP NC
il a déjà fondu le gros morceau de sucre
(uses rule 21.25 GV -> V SNP)

PF VA VN AD AN NC
elle a glissé cette premiere fois
(uses rule 21.25 GV -> V SNP)

PF VA VN DN NC
il a fondu le sucre
(uses rule 21.25 GV -> V SNP)

PF VA VN DN AQ NC
il a grossi le petit chien
(uses rule 21.25 GV -> V SNP)

PF VC EP IN NC EP DN NC
ils sont dans une cage dans les zoos
(uses rule 10.2 SN** -> EP NP* EP NP*)

PF VD DG VN DN AQ NC
elle peut pas avancer la grande locomotive
(uses rule 21.25 GV -> V SNP)

PF VD DV VT NP
elle peut bien attendre Madeleine (said for "Madeleine peut
bien attendre)
(uses rule 21.25 GV -> V SNP)

PF VD PS VT DN NC
je vais les écraser la voiture
(uses rule 22.2 GV' -> PS V SNP)

```
PF VD PS VT PI EP DN NC          tout à  1'heure
je vais la demander tout à  1'heure
(uses rule 9.3 SN* -> NP* EP NP*)

PF VD PS PS VT                   stuttering
je vais te t'attaquer
(uses rule 22.12 GV' -> PS PS V)

PF VD PS VT IN NC EP NP
je vais lui donner une fessée à Michel
(uses rule 22.2 GV' -> PS V SNP and
(9.3 SN* -> NP* EP NP*)

PF VD PS VT AO NC EP DN NC
je vais lui foutre mon panier dans le chauffe-eau
(uses rule 22.2 GV' -> PS V SNP and
(9.3 SN* -> NP* EP NP*)

PF VD VN DN NC
elle peut passer la voiture
(uses rule 21.25 GV -> V SNP)

PF VD VT AO PI AO NC             stuttering
je vais manger mon tout tout mon croissant
(uses rule 21.34 GV -> V SNP SNS)

PF VD VT EP NC AQ EP NC AQ
je vais téléphoner au bras droit au bras gauche
(uses rule 21.35 GV -> V SNP SN**)

PF VD VT PT                      should be PF
je vais boire moi
je vais écrire moi
(uses rule 21.27 GV -> V PI)

PF VN AO NC PF PF                repetition
je roule mon pull moi moi
(uses rule 21.34 GV -> V SNP SNS)

PF VN CS PF VT PT                should be PF
tu vois comment je fais moi
(uses rule 21.27 GV -> V PI)

PF VN DN NC KO
elle glisse la voiture boum
(uses rule 21.25 GV -> V SNP)

PF VN DV DN NC
elle roule là la voiture
(uses rule 21.28 GV -> V DV SNP)
```

```
PF VN EP NC EP DN NC
ils jouaient du pianon de l'accordéon
(uses rule 21.35 GV -> V SNP SN**)

PF VN EP NC EP NC EP NC
il donne des cravates aux gens aux monsieurs
(uses rule 21.35 GV -> V SNP SN**)

PF VN IN NC IN NC AQ
je prends un crayon un crayon noir
(uses rule 21.34 GV -> V SNP SNS)

PF VT AC NC NC
ils avaient deux stylo bille
(uses rule 21.34 GV -> V SNP SNS)

PF VT DG CS PF PS VN NP
il faut pas qu'elle me sorte Ginette
(uses rule 22.2 GV' -> PS V SNP)

PF VT DG PD CS PD VC EP NP PS VR DG
je sais pas ce que c'est à  Porquerolles me souviens pas
(uses rule 5.3 S' -> SNS SV CS SNS SV and 21.20 GV-> COP SN GV')

PF VT DG PD CS PD VC PD                    should be PD PR PD
je sais pas ce que c'est ça

PF VI DG EP NC
je sais pas du blé
(uses rule 21.29 GV -> V DG SNP)

PF VT DV NP
il avait encore Pascal (dit pour "qu'avait encore Pascal?")
(uses rule 21.28 GV -> V DV SNP)

PF VT NC NC
il fait nuit jour
(uses rule 21.34 GV -> V SNP SNS)

PF VT NP                          should have two derivations
tu manges Papa
(uses rule 21.25 GV -> V SNP)

PF VI                             incomplete
il a

PF VT IN NC EP EP DN NC           stuttering
J'ai un pistolet dans dans le pantalon
(uses rule 9.5 SN* -> NP* EP EP NP*)
```

PF VA VN DN NC IN NC
J'ai vu les animaux un zoo
(uses rule 21.34 GV -> V SNP SNS)

PF VA VN EP EP DN NC stuttering
j'ai joué avec avec l'autocar
(uses rule 10.5 SN** -> EP EP NP*)

PF VA VN EP EP NC stuttering
j'ai joué au au loup
(uses rule 10.5 SN** -> EP EP NP*)

PF VA VN PD EP NC
j'ai pris celui-là de mouchoir
(uses rule 21.35 GV -> V SNP SN**)

PF VA VT IN AQ NC EP EP NC NC stuttering
j'ai fait une grande maison avec des nallumettes allumettes
(uses rule 21.24 GV -> V SNP SNS)

PF VD DG VT
tu fais pas étonnant (for "ce que tu fais est pas étonnant")
(uses rule 18.10 SV -> AUX2 DG GV)

PF VT DG PI EP NC CS PF PS DV VI tout de suite
je veux pas tout de suite que tu t'en vas
(uses rule 9.3 SN* -> NP* EP NP*)

PF VT DN NC
il fait le monsieur (for "que fait le monsieur?")
(uses rule 21.25 GV -> V SNP)

PF VD DG PS PS VN stuttering
je veux plus te te voir
(uses rule 22.12 GV' -> PS PS V)

PF VT NP
tu fais Madeleine (for "que fais-tu Madeleine?")
(uses rule 21.25 GV -> V SNP)

PF VT CS PF PS VD VI IN NC
je veux que tu le fasses marcher un coup
(uses rule 21.54 GV -> V' SNS)

PF VD VI CS PD VI
tu peux rester comme ça viens
(uses rule 22.3 S' -> SNS SV CS SNS SV)

PF VC EP NC EP VI NP should have 2 derivations
tu es en train d'écrire Madeleine

```
(for "qu'écris-tu Madeleine?")
(uses rule 22.17 GV' -> EP V SNP)

PF VI IN NC PR VN PT
t'auras un bateau qui roule toi
(uses rule 21.37 GV -> V PT)

PF VA VN PT
t'as fini toi
(uses rule 21.37 GV -> V PT)

PF VC DV NC                          should be AQ
moi je suis encore malade
(no rule VC DV AQ)

PF PF VA DG VN DG DV                 should be DG DG
moi j'avais pas vu non plus

PF PS VT DV PI EP PT                 should be DV DV instead of DV PI
elle m'amène quand même chez elle

PF VN AO NC DN NC EP NC              stuttering
elle vole ses sous les sous du monsieur
(uses rule 21.34 GV -> V SNP SNS)

PI PF VA VN
rien j'ai vu
(uses rule 7.3 SNS -> SN* PF)

PI VC PI  DN AC
on est tous les deux
(no rule NP* -> AI DET AC)
(uses rule 21.15 GV -> COP SN SNS)

PI VD VT PI DN AC
on va écrire tous les deux
(uses rule 21.34 GV -> V SNP SNS)

PI EP DN NC PF PS VT EP NC           should be AI EP DN NC
tout à l'heure t'en veux des croissants
(uses rule 7.3 SNS -> SN* PF and 9.3 SN* -> NP* EP NP*

PI EP DN NC VT EP NC                 should be AI EP DN NC
tout à l'heure a des enfants

PI EP NC                             should be AI EP NC
tout de suite
(uses rule 12.3 SN*1 -> NP* EP NP*)

PI EP DN NC
```

```
tout en l'air
(uses rule 12.3 SN*1 -> NP* EP NP*)

PI VT DN NC EP NP NP
tout mangé le croissant de maman papa
(uses rule 5.2 S' -> SNS SV)

PR PY VT DV NP                    there should have 2 derivations
qu'y a dehors Madeleine
(rule 21.28 GV -> PY DV SNP)

PR PD VA VD VI
qui c'est va tomber
(uses rule 19.2  AUX1 GV -> VA VD GV)

PR LC PF PS VA VT
qu'est-ce il l'a acheté

PR LC PR PF VD PS VD VT          stuttering
qu'est-ce elle fait lui fait faire
(uses rule 18.8 SV -> AUX2 GV)

PR LC PF VD VT NP
qu'est-ce elle fait faire Mamie?
(uses rule 21.25 GV -> V SNP)

PR LC PF VT AD NC AD
qu'est-ce il fait ce garçon là?
(uses rule 21.25 GV -> V SNP)

PR LC PF VT DV DN NC
qu'est-ce il fait là le veau?
(uses rule 21.28 GV -> V DV SNP)

PR LC PF VT NP DV
qu'est-ce elle fait Mamie là?
qu'est-ce il fait Michel là?
(rule 21.26 GV -> V SNP DV)

PS VA VI
l'est tombé
(l' should have been coded DN,PF,PS)

PS VN
la cassé

PS VN NP
la cassé Madeleine
(uses rule 18.2 GV' -> PS V SNP)
```

```
PY VT EP NC EP NC
avait du moteur du bruit
(uses rule 10.2  SN** -> EP NP* EP NP*)

VA VN PT
a fini moi
(uses rule 21.37 GV -> V PT)

VA VT IN NC
a a une poire    stuttering

VD PS VI DN NC  stuttering
fait la fait la cigarette

VD VN PT        should be PF
veux tourner moi

VD VT EP NP PR PS PS VN IN NC
faut dire à Ginette qu'elle me donne un suppositoire
(uses rule 21.52 GV -> V SNP PR S')

VD VT DN AQ NP
va chercher le gros Philippe
(uses rule 11.12 NP* -> DET ADJP NP)

VI PI EP DN NC
reviendra tout à l'heure
(uses rule 21.54 GV -> V' SNS)

VN DG EP NC                        should be AQ
donne pas des bleus

VN DV DN NC
roule bien le camion
(uses rule 21.25 GV -> V SNP)

VN PT                    should be PF
travailler moi
(uses rule 21.37 GV -> V PT)

VT AD NC EP DN NC
faisait cette dame à l'ecole
(uses rule 21.35 GV -> V SNP SNS)

VT EP NP AO NC
oubliée chez Misette ma montre
(uses rule 21.34 GV -> V SNP SNS)

VT PD PR PF VA VT NP            should have two derivations
regarde ce que j'ai fait Madeleine
```

(uses rule 21.25 GV -> V SNP)

VT EP AQ
manger du seul
(uses rule 21.51 GV -> V EP AQ)

VT EP NP EP DN NC
connais chez Misette dans la boîte
(uses rule 21.35 GV -> V SNP SN**)

VT NP PR PF VA VT DN AQ NC EP DN AQ NC wrong derivation
regarde Marc que j'ai met la grosse caravane
à la petite voiture
(uses rule 21.52 GV -> V SNP PR S')

VT NP PR PF VT
regarde Madeleine que je fais
(uses rule 21.52 GV -> V SNP PR S')

VT PT DN NC EP NP should be PF
écris toi la main de Philippe
(uses rule 21.39 GV -> V PT SNP)

VT PT NP
mettre toi Madeleine
(uses rule 21.39 GV -> V PT SNP)

VT PI AQ NP should be DM AQ
peigner tout seul Philippe
(uses rule 21.34 GV -> V SNP SNS)

VT PR LC PY VA DV VT repetition
regarde qu'est-ce y a dedans devine
(uses rule 21.79 GV -> V PR LC PY SV)

VT DG EP NC should be AQ
a plus de roses

VT IN AI NC IN AI NC repetition
a une autre histoire une autre histoire
(uses rule 21.34 GV -> V SNP SNS)

Appendix II

Incorrect Probabilistic Disambiguation

Table IX below presents the utterance types, together with some examples, that were judged to have been incorrectly disambiguated by the probabilitic disambiguation process. We should note that some incorrect analyses result from the first parse, where only one lexical form is recognized and that recognition is incorrect. Only incorrect analyses from the probabilistic disambiguation are counted into the calculations in Table VIII. Such cases are denoted ***(0). For example, we do not include in our statistics the utterance tu peux rester comme ca, viens which is coded PF VD VI,VC CS PD VI and recognized as PF VD VC CS PD VI, since it is the first parsing which analysed incorrectly this utterance. We signal such cases by ***(0).

In other cases, for the same string, one utterance will be correctly analysed, while one other will not. For example, the string DV,EP AQ,NC corresponds to the two following utterances: en bas and en plastique. The string EP NC describes correctly en plastique, but not en bas. Such cases are signaled by ***(1), which means that we count one incorrect disambiguation.

Out of 660 ambiguous types, 88 were solved in an unsatisfying way (out of 938 tokens, 133 were solved in an unsatisfying way). This corresponds to 13.33 percent of the types and 14.17 percent of the tokens. (There are 22 types (66 tokens) solved in an unsatisfying way

when one takes into account utterances down to frequency 2). The incorrect disambiguations fall mainly into the following six cases:

- Cases where DV is chosen instead of DG or DM:

  Ex: PS VT DV,DG is disambiguated as PS VT DV.

  "en a plus".

  In this case, it is only the phonetic form which enables the listener to understand whether plus means no more or some more.

  There are 16 types (39 tokens) with DV,DG or DG,DM

- Cases where the common noun is chosen instead of the verb:

  Ex: PF VC,VA NC,VN DV is disambiguated as PF VC NC DV:

  "elle est rentrée dedans"

- Cases where the common noun is chosen instead of the adjective:

  Ex:DG AQ,NC is disambiguated as DG NC: "pas malade"

  - Cases where PR is chosen instead of CS:

  Ex: DV PF VC DN NC PI,AI NC CS,PR PF,PS VA VT is disambiguated as DV PF VC DN AI NC CS PF VA VT:

  "où ils sont les autres livres que t'as achetés?"

- Cases where VC is chosen rather than VI:

  Ex: PF VC,VI EP DN NC is disambiguated as PF VC EP DN NC:

  "je veux rester à la maison"

- Cases where NC is chosen instead of PI

  Ex: NC,PI is disambiguated as NC: "personne".

TABLE IX

Incorrect Disambiguations

Frequency Original type	Incorrect disambiguation Example(s)
6 PS,DV,EP VT DV,DG	PS VT DV should be DG en a plus
6 PS,DV,EP VT DV,DG EP NC	PS VT DV EP should be DG en a plus des roses en a plus de roses en a plus d'école en a plus de croissant
5 DV,DG NC	DV NC should be DG plus lait plus pneu plus musique
5 PF PS,EP VT DV,DG	PF PS VT DV should be DG j'en veux plus
4 PR,DV PD VC DV	PR PD VC DV should be DV où c'est la
3 DG PF PS,EP VT DV,DG	DG PF PS VT DV should be DG non j'en veux plus
3 DV,DG,DM DV	DV DV should be DG plus là [notice however that in the case of "plus près" it should be DV, or DM]
3 DV,DG EP NC ***(0)	DV EP NC plus de chocolat [ambiguous]
3 NP VC,VI DV	NP VC DV should be VI Maman reste là
3 PF PS VT CS NC,PI PS VT	PF PS VT CS NC PS VT preferably PI je le cache puisque personne le sait

3 PI,DM DV	PI DV tout bêtement	should be DM
3 PS,DN VN,PO	PS VN la tienne	should be DN PO
3 VN NC,DV	VN NC monter dessus	should be DV
3 VT NC,PI	VT NC a personne	preferably PI
2 AQ,NC,KO	NC bon	should be AQ or KO
2 DG NC,PI	DG NC non personne	preferably PI
2 DV EP NC,NP CC NP	DV EP NC CC NP voilà pour Monsieur et Madame	preferably NP
2 DV,EP AQ,NC ***(1)	EP NC en bas en plastique	should be AQ is correct
2 NC,DV,LC EP DN NC	DV EP DN NC mal à la tête	should be NC
2 NC,PI	NC personne	preferably PI
2 PF VC NC,DV	PF VC NC il est dessous	should be DV
2 PF VC,VA AQ,VT DN NC	PF VA VT DN NC il est vide le pot elle est vide la bouteille	should be VC AQ
2 PY PS VT DV,DG	PY PS VT DV y en a plus	should be DG

Frequency 1

AC NC PF,PS VA VT	AC NC PS VA VT trois heures t'avais mis	should be PF
AQ,NC,DV,LC EP NC	DV EP NC mal au ventre	should be NC

```
CC DV AI DN NC VN NC,DV          CC DV AI DN NC VN NC    should be DV
                                 et puis tout le monde monte dessus

CC,DV PD VC AO NC CS             DV PD VC AO NC CS PF VA VT LC
PF VA VT LC                      mais c'est ton fou que j'ai mis
***(0)                           là-bas
                                 [here it is at the level of tree-1 that
                                 PR has been eliminated]

DV AQ,NC                         DV NC           should be AQ
                                 si malade

CS,DV PF VA DG VT NC             DV PF VA DG VT NC AD NC AD
AD NC AD                         pourquoi il a pas mis son ce
***(0)                           ce manteau-là
                                 [Philippe stuttered. "son"
                                 was coded as NC,AO.
                                 The grammar did recognize NC]

CS,DV PF VN CS PF VD             DV PF VN CS PF VD EP VT EP NC
EP VT EP NC EP DN NC             EP DN NC
***(0)                           comment tu ouvres comment tu fais
                                 pour mettre du feu sur la cigarette
                                 [Philippe stuttered]

DG NC,VT                         DG NC           ambiguous
                                 pas montre

DG PF VC,VA NC,VN DV             DG PF VC NC DV  should be VA VN
                                 non elle était sortie aussi

DG PF VD VC,VI DV                DG PF VD VC DV  Should be VD VI
                                 non je veux rester ici

DN AQ NC,NP PF VC DG AQ          DN AQ NC PF VC DG AQ CS PI PF
CS PI PF,PS VN PI,AI AO NC       VN AI AO NC      should be PI PS
                                 le méchant monsieur il était pas
                                 pas content qu'on lui prende
                                 toutes ses sous

DN NC AQ PF PS VT NC,DV          DN NC AQ PF PS VT DV    should be NC
                                 le pied droit il me fait mal

DN NC PF VC DV,EP NC             DN NC PF VC EP NC
                                 ["dessous" coded as NC,EP is parsed
                                 as NC]
                                 le moteur il est en dessous

DN NC PF VT PD PR,CS PF VT       DN NC PF VT PD CS PF VT should be PR
```

	le chat il regarde ce qu'il fait
DN NC PR VC,VA AQ,VN NC,DV	DN NC PR VC AQ NC should be DV la voiture qui est cassée derrière
DV PF PS VT DV,DG	DV PF PS VT DV should be DG maintenant je le retrouve plus
DV PF VC DN PI,AI NC CS,PR PF,PS VA VT	DV PF VC DN AI NC CS PF VA VT should be PR où ils sont les autres livres que t'as achetés
DV PF VC DV,DG EP DN NC	DV PF VC DV EP DN NC should be DG là elle est plus sur les rails
DV PF VC PD DN NP EP AD NC,NP ***(0)	DV PF VC PD DN NP EP AD NC où il est ce le gronomme de ce monsieur [Philippe stuttered]
DV PI PF,PS VA VT IN NC ***(0)	DV PI PF VN DV AD NC là on il travaille aussi ce monsieur [Philippe stuttered]
DV VT IN PI PI,DV ***(0)	DV VT IN PI DV the first DV should be PS en a un autre même
DV,CS PD PD VC	DV PD PD VC que ce c'est
DV,CS DN AQ NC	DV DN AQ NC que le vieux monsieur
DV,CS PD PD VC PD NP	DV PD PD VC PD NP should be PR que ce c'est ça maman
DV,CS PF PF,PS VT PI,AI NC	DV PF PS VT AI NC should be CS que je t'écrise quelque chose
DV,CS PF PS VI	DV PF PS VT should be CS que je te mette
DV,CS PF VA VT	DV PF VA VT que j'ai fait
DV,CS PF VD VI PR	DV PF VD VT PR should be CS qu'elle vienne chercher qui

DV,CS PF VT DV DN AQ NC	DV PF VT DV DN AQ NC should be PR qu'elle fait là la petite fille
DV,CS PI VT	DV PI VT should be PR or CS qu'on fait
DV,DG PI	DV PI should be DG plus rien
DV,EP,PS EP,PS NC ***(0)	DV EP NC en en bois [Philippe stuttered]
DV,EP,PS NC,VN	EP NC should be EP VN en courant
EP NC,DV	EP NC should be EP DV par dessus
EP PR,CS AO NC VI	EP PR AO NC VI should be CS pour que son heure vienne
EP PR,CS PF VI DG	EP PR PF VI DG should be CS pour qu'ils tombent pas
EP PR,DV	EP PR should be DV par où
EP VN NC,DV	EP VN NC should be DV pour monter dessus
EP,DV DN NC PF VD VN NC,DV	EP DN NC PF VD VN NC should be DV and DV après le monsieur il va tomber dessus
KO EP PR,CS DN NC PF VI DG	KO EP PR DN NC PF VI DG should be CS ben pour que le train il déraille pas
KO PF VA VT DM,DV AQ,NC ***(0)	KO PF VA VT DM AQ should be DM NC but this is not generated by the grammar il a eu trop froid
LC CS PF VC DG,DM AQ ***(0)	LC CS PF VC DG AQ parce qu'il est plus beau
LC CS PF VC DG,DM AQ DN NC	LC CS PF VC DG AQ DN NC should be DM AQ parce qu'il est plus joli le train

LC CS PF VT CS PF VI,VC DV	LC CS PF VT CS PF VC DV should be VI parce que je veux que tu restes ici
LC CS PF VT NC,DV,LC	LC CS PF VT NC should be DV parce qu'elle dessine mal
NC,DV DV	DV DV should be NC mal partout
NC,KO DV PF VI DG ***(0)	KO DV PF VI DG attention qu'elle tombe pas here it is at tree-1 that DV has been chosen
NC,PI PR VN	NC PR VN should be PI personne qui parle
NC,VT DV ***(0)	NC DV montre là [ambiguous]
NP PF VI,VC DV	NP PF VC DV should be VI Madeleine elle reste ici
PD PD VC DG,DM AQ	PD PD VC DG AQ probably DM ça c'est plus difficile
PD PY VT NC,DV	PD PY VT NC should be DV c'y a dessus
PD VC DV CS PF PS VA VT AD,PD NC ***(0)	PD VC DV CS IT PS VA VT AD NC c'est où que tu l'as acheté ce balai
PD VC PR,DV	PD VC PR should be DV c'est où
PD VC PR VC AQ,NC DV	PD VC PR VC NC DV should be AQ c'est quoi qui est rose là
PF VA VN EP,DV NC,EP DN NC	PF VA VN DV EP DN NC il est passé en dessous le cerf
PF PF,PS VT DN NC EP DN NC	PF PS VT DN NC EP DN NC should be PF toi t'as la tête à l'envers
PF VD DG VN NC,DV	PF VD DG VN NC should be DV je peux pas appuyer dessus

PF VD VC,VI PI,AI DN NC EP NP	PF VD VC AI DN NC EP NP should be VI je vais rester tout le temps chez Mamie
PF VD VC,VI CS PD VI ***(0)	PF VD VI CS PD VI tu peux rester comme ça, viens [at tree-1, DV should have been parsed instead of CS]
PF VD VI,VC EP DN NC	PF VD VC EP DN NC should be VI je veux rester à la maison
PF VD VI,VC EP DN NC EP NP	PF VD VC EP DN NC EP NP should be VI je veux rester sans les parents chez Mamie
PF VD VT DN NC PR,CS NP PF PS VA VN	PF VD VT DN NC CS NP PF PS VA VN should be PR je vais chercher le camion que Minou elle m'a donné
PF VC,VI DV	PF VC DV should be VI tu resteras ici
PF VT DG PD CS,PR PF,PS VA VT	PF VT DG CS PF VA VT should be PR j'entends pas ce que t'as dit
PF VT NC,DV,LC	PF VT NC should be DV je mange mal
PF,PS PS,EP VT DV,DG ***(0)	PF PS VT DV ambiguous t'en veux plus
PF,PS VA VN PT ***(0)	PF VA VN PT t'as fini toi [at tree-1 PF has been discarded]
PI PF,PS VT DN NC EP NC	PI PF VT DN NC EP NC should be PS on nous montre la queue du chat
PI VN NC,DV	PI VN NC should be DV on tire dessus
PR LC PY VA VT NC,DV	PR LC PY VA VT NC should be DV qu'est-ce y a dessiné dessus
PR,DV PD VC	PR PD VC should be DV où c'est

PR,DV PD VC DN NC EP DN NC	PR PD VC DN NC EP DN NC should be DV où c'est le derrière de l'hélicoptère
PS,DV VT DV,DG EP AQ NC	PS VT DV EP AQ NC should be DG en a plus des petits trous
PS,DV VT DV,DG NP	PS VT DV NP should be DG en a plus Madeleine
PS,DV,EP VT EP PI EP NC	DV VT EP PI EP NC should be PS en a d'autres par terre
PY PS VT DV,DG NP	PY PS VT DV NP should be DG y en a plus Papa
PY VT DV,DG NC,NP	PY VT DV NC should be DG y a plus monsieur
VT NC,PI DV DV	VT NC DV DV should be PI a personne aujourd'hui
VT NC,PI PR PS VA VT IN AQ NC	VT NC PR PS VA VT IN AQ NC should be PI a personne qui m'a achete un petit moulin
VT NC,PI PR VI	VT NC PR VI should be PI a personne qui vient

References

1. Dubois, J., Grammaire structurale du français: nom et pronom. Paris: Larousse, 1970.

2. Gammon, E. M., A syntactical analysis of some first-grade readers. Technical Report No. 155, June 22, 1970, Stanford University, Institute for Mathematical Studies in the Social Sciences.

3. Grevisse, M., Le bon usage. Grammaire française avec des remarques sur la langue francaise d'aujourd'hui. Gemblouz, France: Duculot, 1969.

4. Martinet, M. A., De l'économie des formes du verbe en francais parlé. A. G. Hatcher & K. L. Selid (Eds.), Studia philologica et litteraria in honorem L. Spitzer. Bern, Switzerland, Francke, 1958.

5. Robert, P., Le petit Robert. Paris: Societé du nouveau Littré, 1967.

6. Smith, N.W., A question-answering system for elementary mathematics. Technical Report No. 227, April 19, 1974, Stanford University, Institute for Mathematical Studies in the Social Sciences.

7. Smith, R. L., Jr., The syntax and semantics of ERICA. Technical Report No. 185, June 14, 1972, Stanford University, Institute for Mathematical Studies in the Social Sciences.

8. Suppes, P., Probabilistic grammars for natural languages. Technical Report No. 154, May 15, 1970, Stanford University, Institute for Mathematical Studies in the Social Sciences.

9. Suppes, P., Smith, R., Léveillé, M., The French syntax and semantics of Philippe, part 1: noun phrases, Technical

Report No. 195, November 3, 1972, Stanford University, Institute for Mathematical Studies in the Social Sciences.

(Continued from inside front cover)

165 L. J. Hubert. A formal model for the perceptual processing of geometric configurations. February 19, 1971. (A statistical method for investigating the perceptual confusions among geometric configurations. Journal of Mathematical Psychology, 1972, 9, 389-403.)

166 J. F. Juola, I. S. Fischler, C. T. Wood, and R. C. Atkinson. Recognition time for information stored in long-term memory. (Perception and Psychophysics, 1971, 10, 8-14.)

167 R. L. Klatzky and R. C. Atkinson. Specialization of the cerebral hemispheres in scanning for information in short-term memory. (Perception and Psychophysics, 1971, 10, 335-338.)

168 J. D. Fletcher and R. C. Atkinson. An evaluation of the Stanford CAI program in initial reading (grades K through 3). March 12, 1971. (Evaluation of the Stanford CAI program in initial reading. Journal of Educational Psychology, 1972, 63, 597-602.)

169 J. F. Juola and R. C. Atkinson. Memory scanning for words versus categories. (Journal of Verbal Learning and Verbal Behavior, 1971, 10, 522-527.)

170 I. S. Fischler and J. F. Juola. Effects of repeated tests on recognition time for information in long-term memory. (Journal of Experimental Psychology, 1971, 91, 54-58.)

171 P. Suppes. Semantics of context-free fragments of natural languages. March 30, 1971. (In K. J. J. Hintikka, J. M. E. Moravcsik, and P. Suppes (Eds.), Approaches to natural language. Dordrecht: Reidel, 1973. Pp. 221-242.)

172 J. Friend. INSTRUCT coders' manual. May 1, 1971.

173 R. C. Atkinson and R. M. Shiffrin. The control processes of short-term memory. April 19, 1971. (The control of short-term memory. Scientific American, 1971, 224, 82-90.)

174 P. Suppes. Computer-assisted instruction at Stanford. May 19, 1971. (In Man and computer. Proceedings of international conference, Bordeaux, 1970. Basel: Karger, 1972. Pp. 298-330.)

175 D. Jamison, J. D. Fletcher, P. Suppes, and R. C. Atkinson. Cost and performance of computer-assisted instruction for education of disadvantaged children. July, 1971.

176 J. Offir. Some mathematical models of individual differences in learning and performance. June 28, 1971. (Stochastic learning models with distribution of parameters. Journal of Mathematical Psychology, 1972, 9(4),)

177 R. C. Atkinson and J. F. Juola. Factors influencing speed and accuracy of word recognition. August 12, 1971. (In S. Kornblum (Ed.), Attention and performance IV. New York: Academic Press, 1973.)

178 P. Suppes, A. Goldberg, G. Kanz, B. Searle, and C. Stauffer. Teacher's handbook for CAI courses. September 1, 1971.

179 A. Goldberg. A generalized instructional system for elementary mathematical logic. October 11, 1971.

180 M. Jerman. Instruction in problem solving and an analysis of structural variables that contribute to problem-solving difficulty. November 12, 1971. (Individualized instruction in problem solving in elementary mathematics. Journal for Research in Mathematics Education, 1973, 4, 6-19.)

181 P. Suppes. On the grammar and model-theoretic semantics of children's noun phrases. November 29, 1971.

182 G. Kreisel. Five notes on the application of proof theory to computer science. December 10, 1971.

183 J. M. Moloney. An investigation of college student performance on a logic curriculum in a computer-assisted instruction setting. January 28, 1972.

184 J. E. Friend, J. D. Fletcher, and R. C. Atkinson. Student performance in computer-assisted instruction in programming. May 10, 1972.

185 R. L. Smith, Jr. The syntax and semantics of ERICA. June 14, 1972.

186 A. Goldberg and P. Suppes. A computer-assisted instruction program for exercises on finding axioms. June 23, 1972. (Educational Studies in Mathematics, 1972, 4, 429-449.)

187 R. C. Atkinson. Ingredients for a theory of instruction. June 26, 1972. (American Psychologist, 1972, 27, 921-931.)

188 J. D. Bonvillian and V. R. Charrow. Psycholinguistic implications of deafness: A review. July 14, 1972.

189 P. Arabie and S. A. Boorman. Multidimensional scaling of measures of distance between partitions. July 26, 1972. (Journal of Mathematical Psychology, 1973, 10,)

190 J. Ball and D. Jamison. Computer-assisted instruction for dispersed populations: System cost models. September 15, 1972. (Instructional Science, 1973, 1, 469-501.)

191 W. R. Sanders and J. R. Ball. Logic documentation standard for the Institute for Mathematical Studies in the Social Sciences. October 4, 1972.

192 M. T. Kane. Variability in the proof behavior of college students in a CAI course in logic as a function of problem characteristics. October 6, 1972.

193 P. Suppes. Facts and fantasies of education. October 18, 1972. (In M. C. Wittrock (Ed.), Changing education: Alternatives from educational research. Englewood Cliffs, N. J.: Prentice-Hall, 1973. Pp. 6-45.)

194 R. C. Atkinson and J. F. Juola. Search and decision processes in recognition memory. October 27, 1972.

195 P. Suppes, R. Smith, and M. Léveillé. The French syntax and semantics of PHILIPPE, part 1: Noun phrases. November 3, 1972.

196 D. Jamison, P. Suppes, and S. Wells. The effectiveness of alternative instructional methods: A survey. November, 1972.

197 P. Suppes. A survey of cognition in handicapped children. December 29, 1972.

198 B. Searle, P. Lorton, Jr., A. Goldberg, P. Suppes, N. Ledet, and C. Jones. Computer-assisted instruction program: Tennessee State University. February 14, 1973.

199 D. R. Levine. Computer-based analytic grading for German grammar instruction. March 16, 1973.

200 P. Suppes, J. D. Fletcher, M. Zanotti, P. V. Lorton, Jr., and B. W. Searle. Evaluation of computer-assisted instruction in elementary mathematics for hearing-impaired students. March 17, 1973.

201 G. A. Huff. Geometry and formal linguistics. April 27, 1973.

202 C. Jensema. Useful techniques for applying latent trait mental-test theory. May 9, 1973.

203 A. Goldberg. Computer-assisted instruction: The application of theorem-proving to adaptive response analysis. May 25, 1973.

204 R. C. Atkinson, D. J. Herrmann, and K. T. Wescourt. Search processes in recognition memory. June 8, 1973.

205 J. Van Campen. A computer-based introduction to the morphology of Old Church Slavonic. June 18, 1973.

206 R. B. Kimball. Self-optimizing computer-assisted tutoring: Theory and practice. June 25, 1973.

207 R. C. Atkinson, J. D. Fletcher, E. J. Lindsay, J. O. Campbell, and A. Barr. Computer-assisted instruction in initial reading. July 9, 1973.

208 V. R. Charrow and J. D. Fletcher. English as the second language of deaf students. July 20, 1973.

209 J. A. Paulson. An evaluation of instructional strategies in a simple learning situation. July 30, 1973.

210 N. Martin. Convergence properties of a class of probabilistic adaptive schemes called sequential reproductive plans. July 31, 1973.

(Continued from inside back cover)

211 J. Friend. Computer-assisted instruction in programming: A curriculum description. July 31, 1973.

212 S. A. Weyer. Fingerspelling by computer. August 17, 1973.

213 B. W. Searle, P. Lorton, Jr., and P. Suppes. Structural variables affecting CAI performance on arithmetic word problems of disadvantaged and deaf students. September 4, 1973.

214 P. Suppes, J. D. Fletcher, and M. Zanotti. Models of individual trajectories in computer-assisted instruction for deaf students. October 31, 1973.

215 J. D. Fletcher and M. H. Beard. Computer-assisted instruction in language arts for hearing-impaired students. October 31, 1973.

216 J. D. Fletcher. Transfer from alternative presentations of spelling patterns in initial reading. September 28, 1973.

217 P. Suppes, J. D. Fletcher, and M. Zanotti. Performance models of American Indian students on computer-assisted instruction in elementary mathematics. October 31, 1973.

218 J. Fiksel. A network-of-automata model for question-answering in semantic memory. October 31, 1973.

219 P. Suppes. The concept of obligation in the context of decision theory. (In J. Leach, R. Butts, and G. Pearce (Eds.), Science, decision and value. (Proceedings of the fifth University of Western Ontario philosophy colloquium, 1969.) Dordrecht: Reidel, 1973. Pp. 1-14.)

220 F. L. Rawson. Set-theoretical semantics for elementary mathematical language. November 7, 1973.

221 R. Schupbach. Toward a computer-based course in the history of the Russian literary language. December 31, 1973.

222 M. Beard, P. Lorton, B. W. Searle, and R. C. Atkinson. Comparison of student performance and attitude under three lesson-selection strategies in computer-assisted instruction. December 31, 1973.

223 D. G. Danforth, D. R. Rogosa, and P. Suppes. Learning models for real-time speech recognition. January 15, 1974.

224 M.R. Raugh and R. C. Atkinson. A mnemonic method for the acquisition of a second-language vocabulary. March 15, 1974.

225 P. Suppes. The semantics of children's language. (American Psychologist, 1974, 29, 103-114.)

226 P. Suppes and E. M. Gammon. Grammar and semantics of some six-year-old black children's noun phrases.

227 N. W. Smith. A question-answering system for elementary mathematics. April 19, 1974.

228 A. Barr, M. Beard, and R. C. Atkinson. A rationale and description of the BASIC instructional program. April 22, 1974.

229 P. Suppes. Congruence of meaning. (Proceedings and Addresses of the American Philosophical Association, 1973, 46, 21-38.)

230 P. Suppes. New foundations of objective probability: Axioms for propensities. (In P. Suppes, L. Henkin, Gr. C. Moisil, and A. Joja (Eds.), Logic, methodology, and philosophy of science IV: Proceedings of the fourth international congress for logic, methodology and philosophy of science, Bucharest, 1971. Amsterdam: North-Holland, 1973. Pp.515-529.)

231 P. Suppes. The structure of theories and the analysis of data. (In F. Suppe (Ed.), The structure of scientific theories. Urbana, Ill.: University of Illinois Press, 1974. Pp. 267-283.)

232 P. Suppes. Popper's analysis of probability in quantum mechanics. (In P. A. Schilp, (Ed.), The philosophy of Karl Popper. Vol. 2. La Salle, Ill.: Open Court, 1974. Pp.760-774.)

233 P. Suppes. The promise of universal higher education. (In S. Hook, P. Kurtz, and M. Todorovich (Eds.), The idea of a modern university. Buffalo, N. Y.: Prometheus Books, 1974. Pp. 21-32.)

234 P. Suppes. Cognition: A survey. (In J. A. Swets and L. L. Elliott (Eds.), Psychology and the handicapped child. Washington, D. C.: U. S. Government Printing Office, 1974.)

235 P. Suppes. The place of theory in educational research. (Educational Researcher, 1974, 3 (6), 3-10.)

236 V. R. Charrow. Deaf English--An investigation of the written English competence of deaf adolescents. September 30, 1974.

237 R. C. Atkinson and M. R. Raugh. An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. October 4, 1974.

238 R. L. Smith, N. W. Smith, and F. L. Rawson. CONSTRUCT: In search of a theory of meaning. October 25, 1974.

239 A. Goldberg and P. Suppes. Computer-assisted instruction in elementary logic at the university level. November 8, 1974.

240 R. C. Atkinson. Adaptive instructional systems: Some attempts to optimize the learning process. November 20, 1974.

241 P. Suppes and W. Rottmayer. Automata. (In E. C. Carterette and M. P. Friedman (Eds.), Handbook of perception. Vol. 1. Historical and philosophical roots of perception. New York: Academic Press, 1974.)

242 P. Suppes. The essential but implicit role of modal concepts in science. (In R. S. Cohen and M. W. Wartofsky (Eds.), Boston studies in the philosophy of science, Vol. 20, K. F. Schaffner and R. S. Cohen (Eds.), PSA 1972, Proceedings of the 1972 biennial meeting of the Philosophy of Science Association, Synthese Library, Vol. 64. Dordrecht: Reidel, 1974.)

243 P. Suppes, M. Léveillé, and R. L. Smith. Developmental models of a child's French syntax. December 4, 1974.